NUMERICAL MODELING OF BARRED SPIRAL GALAXIES

By

Elizabeth M. Moore

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

1992

In memory of S. K. and J. A. Moore

ACKNOWLEDGMENTS

TABLE OF CONTENTS

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy.

NUMERICAL MODELING OF BARRED SPIRAL GALAXIES

By

ELIZABETH M. MOORE

May, 1992

Chairman: Dr. James H. Hunter
Major Department: Astronomy

A two-component, self-consistent computer code to model spiral galaxies was written
and tested and a method of inducing and controlling bar formation is developed.
This work presents a departure from former modeling work done at the University of
Florida, which depended on the beam scheme, a hydrodynamical code with a number of
limitations. In particular, only the gas component could be modeled, no self-gravitational
forces were included, and the viscosity inherent to the code could not be controlled easily.
These shortcomings are overcome in the new algorithm. Most importantly, an attempt
has been made to keep the models self-consistent. No perturbing potentials are imposed
or required to excite bar and spiral structure.

The code can model both the stellar and the gaseous component of a spiral galaxy.
The stellar component feels only gravitational forces, while the gas component feels
both gravitational and viscous forces. In addition, a halo force can be imposed for the
purpose of stabilizing the disk. The code is a hybrid grid/smooth particle code. The

gravitational forces are calculated on a Cartesian grid using a Fast Fourier Transform, while the gas viscous forces are calculated in a smooth particle manner.

A mechanism for creating warm, featureless, stable disks is developed by taking moments of the collisionless Boltzmann equation. In order to induce and control bar and spiral arm formation, the stabilizing stellar velocity dispersions are reduced in the center of the disk, but maintained in the outer regions. A bar forms naturally in the interior and the rotation of this bar helps maintain sprial structure in the outer gas disk. Realistic-looking spiral features are maintained in the gas component for as long as the models are calculated. A wide variety of bar and spiral structure can be formed by varying the size of the unstable central region, the rate of "turn on" of the heating and the halo mass.

We would like to test the model results by comparing them with observations and so a second part of the thesis consists of observing and reducing 21 cm line data of NGC 1398 and NGC 1784 at the Very Large Array. Low (C/D array) and high (B/C) resolution data were obtained, calibrated and combined to make maps of the integrated column density and mean radial velocity of the neutral hydrogen. In the future, the code will be used to try to model these galaxies.

CHAPTER
INTRODUCTION

Historical Background

The major portion of this dissertation is on the implementation and testing of a self-consistent numerical scheme to model the stellar and gas components of barred spiral galaxies. The goal of this chapter is to place such an endeavor in its proper historical context. A brief history of both numerical and analytical approaches to the study of disk stability and spiral structure will be given, followed by a short discussion of how this work contributes to our understanding of spiral galaxies. The Toomre $n = 0$ infinite disk, used extensively throughout this work, is then introduced. The chapter closes with an outline of the remainder of the dissertation.

The use of numerical codes to simulate model galaxies is not new. The first grid codes were written in the late 1960s (Miller and Prendergast, 1968; Hohl and Hockney, 1969). The smooth particle approach found in the present gas calculations was introduced in the mid 1970s (Lucy, 1977; Sanders, 1977; Gingold and Monaghan, 1977). In the mid 1980s, the first major study using a two-component model was completed. This numerical work was not proceeding in a vacuum; analytical studies of bar and spiral formation in galaxies goes back to the 1920s. Both approaches, analytical and numerical, have limitations and the two must be used together in order to advance our understanding of the complexities posed by spiral galaxies. Numerical techniques can be used to investigate a wide variety of problems; here they will be used to study the stability of

flattened galactic disks and the development of structure in spiral galaxies. Accordingly, a brief history of spiral structure theory and of numerical work on galactic dynamics is given. As disk stability analysis will be discussed extensively in Chapter 3, it is only mentioned in passing in this chapter.

The tasks of explaining the dynamics, stability and structure of galactic disks are not expected to be easy. The most complex galactic structures found, spiral galaxies account for more than two-thirds of the largest and brightest galaxies. Approximately one-third of these contain bars. Their morphologies run the gamut from clean, symmetric, global two-armed grand designs to flocculent galaxies, replete with spurs, feathers, etc. It is not surprising that the fundamentals of the origin and evolution of their spiral structure are not yet completely understood, and it is not certain than any one theory will be able to explain all the variety found among spiral and barred spiral galaxies. However, complex as their structure is, spiral galaxies appear to be neither random nor chaotic. Their internal motions remain ordered and fairly symmetrical and "the dynamics governing the spirals should, therefore, be both interesting and ultimately tractable, if difficult." (Ball, 1984 p. 2).

One of the most influential analytical approaches in the field of spiral structure is the density wave theory. Its roots extend to work done by B. Lindblad in the 1940s, but the modern theory, presented in the mid 1960s, is due to C.C. Lin and F. Shu (1964). They proposed that spiral structure is a wave phenomenon, a quasi-stationary density wave, created and maintained by gravitational forces (the Lin-Shu hypothesis). Despite its many successful observational predictions, the validity of the Lin-Shu hypothesis, as

outlined below, is still the subject of debate. Reviews of this theory are found in Wielen (1974), Toomre (1977) and Shu (1982).

The density wave theory attempts to explain the grand design structure found on a global scale in some of the most prominent and striking spiral galaxies. According to the theory, spiral arms at any moment represent the local maxima of a density wave in the galaxy. The wave is composed of both gas and stars and moves relative to these objects; in the inner region of the galaxy, the spiral wave pattern rotates more slowly than material in the disk, while beyond corotation, the pattern sweeps past the material. It is caused and maintained by purely gravitational effects. The wave propagates in a self-consistent manner due to the corresponding wave perturbation in the galactic gravitational field. It affects both the positions and velocities of the particles it encounters (Wielen, 1974).

The density wave theory as developed thus far, is primarily a linear one, although the term density wave is used in nonlinear contexts. Lin and Shu considered only small perturbations in the gravitational field, due to a nonaxisymmetric distribution of matter, in the form of a rigidly rotating, neutral, tightly wound spiral perturbation imposed on a rotationally symmetrical and stationary spiral galaxy (the WKB approximation). The amplitude of the perturbation is assumed to be small and the wave rotates with constant angular frequency. Differential rotation and velocity dispersions act to stabilize the disk, while gravitational forces drive the instability. The two forms of the solution are either stable, (oscillating) modes or instabilities. Spiral structure is explained by wave solutions (modes). To prevent local instabilities from destroying the wave, Lin and Shu assume that Q, the ratio of the amplitude of the velocity dispersion to the Toomre dispersion

minimum (the minimum velocity dispersion necessary to prevent local axisymmetric instabilities) is 1. A self-consistent procedure is followed to find the density wave phase and amplitude functions which allow the perturbation to be a normal mode of oscillation. For such a normal mode, density enhancements and rarefactions associated with the wave pattern produce just the perturbing gravitational forces required to sustain the shape and motion of the wave. For a tightly wound spiral, the maximum of the density wave is in phase with the minimum of the potential wave.

The density wave theory makes a number of predictions in good accord with observations. The first is that the strength of spiral features is inversely proportional to the magnitude of the velocity dispersions. The smaller the dispersion, the greater the relative amplitude of the density wave, as peculiar velocities tend to smooth local inhomogeneities. Thus, the density wave should be more pronounced in the gas and young stellar components of a galaxy than in the older stellar disk. This is in good agreement with observations.

Density wave theory also predicts the preference of two-armed spirals for grand design galaxies. According to the theory, wave structure exists only between the inner and outer Lindblad resonances, the principal range. For two-armed systems (m = 2), this is typically a large fraction of the disk. For m greater than two, the principal range shrinks to a small portion of the galaxy, thus weakening these modal contributions to the overall spiral structure of a galaxy.

Finally, the Lin-Shu theory nicely explains why star formation and dust lanes occur in the spiral arms. The density wave represents a local density maximum, which is more

pronounced in the gas component. The gas responds so nonlinearly that it piles up in radiating shock waves. The large concentration of material behind the shock leads to an increase in star formation downstream from the shock front (Fujimoto, 1968; Roberts, 1969).

Despite its many successes, the Lin-Shu hypothesis has several shortcomings. First, it is a linear theory while the equations that govern the stellar dynamics of a galaxy—continuity, the equation of motion and the energy equation—are nonlinear. There is no reason to believe that the perturbations in density and potential involved with the spiral pattern should be small (Ball, 1984). Moreover, linear treatment of density waves breaks down at the Lindblad resonances and at corotation. At the resonances, the density response is no longer in phase with the wave in the potential and a nonlinear analysis is required. Furthermore, the amplitude of the density wave is undetermined by linear theory (Wielen, 1974).

Second, the density wave theory does not favor either leading or trailing arms. However, although difficult to determine, observational evidence supports the idea that spiral arms are trailing features. The theory fails to explain these observations.

The most serious blow to the density wave theory, however, came in the 1960s, when it was pointed out, first by Goldreich and Lynden-Bell (1965) but more strenuously and repeatedly by Toomre, that the density waves of Lin's type are not quasi-stationary but have a group velocity. Any packet of density waves should propagate radially, inward from corotation for leading spiral waves and outward for trailing spiral waves. The time required for a wave packet to propagate across a typical galaxy is short, on the order

of a few times 10^8 years (Toomre, 1969). This leads to a complication; the energy of the wave would be transported radially away from corotation, and the packet of spiral density waves (for trailing disturbances) would become more tightly wrapped, winding up in less than 10^9 years. A tightly wound leading disturbance would unwind. Such disturbances cannot be permanently neutral waves without some energy replacement. Therefore, a continuous energetic regeneration of an existing density wave or a frequent generation of new density waves had to be found (Wielen, 1974).

These ideas were shortly expanded to the theory of swing amplification by Goldreich and Lynden-Bell (1965), Julian and Toomre (1966), and Toomre and Zang (Toomre, 1981). Toomre and Zang have made the most thorough investigation of this problem, using linear perturbation theory to make numerical studies on a stellar Mestel disk with a rigid, fixed halo component of varying mass. They found that a leading wave perturbation unwound until it hit corotation, where it was sheared into an open trailing wave, which slowly wound up. An unexpected and remarkable result, however, was that the amplitude of the trailing wave was many times that of the initial leading wave; it had somehow been magnified many times. These results are now explained by swing amplification theory. Swing amplification can be understood physically as follows; for a tightly wound spiral arm, the unwinding rate of the arm is slow, but, at corotation, as it swings from a leading to a trailing wave, this rate maximizes at a value comparable to the average stellar epicyclic frequency, κ. For a short period, the unwinding of the arm and the epicyclic motion of the stars are in the same sense, opposite to the direction of rotation. This temporary near match enhances the gravitational effect of the spiral on the stellar

orbit, and the contribution of the star's gravity to the spiral perturbation, and can lead to rapid growth in the strength of the arm over the interval of about one radian when the arm is most open (Binney and Tremaine, 1987).

Swing amplification also works as a feedback mechanism, solving the winding problem by creating an ongoing supply of density waves. In the absence of an inner Lindblad resonance, waves can propagate into the center of the disk. A trailing wave that propagates into the center will emerge as a leading wave propagating outward. A minor leading disturbance will unwind and swing amplify into a short trailing disturbance, which will propagate through the center and emerge as a short leading disturbance, which then unwinds and is amplified further as it swings into a trailing wave. Swing amplification thus explains the preference for trailing spiral waves and provides a mechanism for the regeneration of spiral structure.

Although extremely useful, linear analytical approaches to understanding spiral structure are somewhat limited. In numerical experimentation, unlike an analytical approach, it is not possible to pick out particular modes believed to be important while ignoring unwanted components of the general solution. "Rather, we get immediately involved in the general solution of the posed problem, i.e., the full spectrum of instabilities and waves show up" (Wielen, 1974 p. 348). N-body codes can study the self-consistent evolution of particles as they move under the long range influences of the galactic gravitational potential. Grid codes utilizing Fast Fourier Transforms have traditionally been the method of choice for simulating collisionless disks (Sellwood, 1987).

The pioneers of numerical studies of stellar disks include Miller and Prendergast (1968) and Hohl and Hockney (1969), who independently wrote the first grid codes. Hohl (1971) carried out the first extensive large-scale investigation of the internal dynamics of galaxies using these codes. He worked on disk stability analysis, first on rigidly rotating disks and then on differentially rotating disks. His models showed that the small scale axisymmetric perturbations predicted by Toomre (1964) and Hunter (1963) were not the only instabilities of significance in a galactic model; nonaxisymmetric, global, bar instabilities were also prominent. He verified the importance of heating to stabilize a cold disk against the fast-growing, short-scale perturbations, as predicted by Toomre, but also discovered the significance of heating well in excess of that predicted to locally stabilize a disk against global nonaxisymmetric instabilities. This work and other numerical work done in the early 1970s provided a new understanding of galactic dynamics.

The first grid codes were used mainly for the study of stellar dynamics. At the same time, work was being done on gaseous disks. Roberts, Roberts and Shu (1975) and others studied the gas response to spiral density wave perturbations. A different approach was taken by Sanders and Huntley (1976). They used the beam scheme, developed by Sanders and Prendergast (1974), to study the response of uniform gas disks with no self-gravitational forces to an imposed central force field with a perturbing potential due to a rigidly rotating bar. Once the gas had settled into a steady state it exhibited a strong trailing spiral pattern. They found a strong gas spiral response even with a relatively weak bar. They concluded that the spiral structure was the consequence of viscosity in the gas disk. Their simulations demonstrated that the formation of spiral arms requires

only the presence of a bar or oval distortion and a dissipative interstellar medium. A spiral density wave in the stellar component is not needed (Sanders and Huntley, 1976).

Although two-component models first appeared in the 1970s (Miller, Prendergast and Quirk, 1970; and Hohl, 1975), it was not until the mid 1980s that a self-consistent study of both components of a disk was completed (Carlberg and Freeman, 1985). They imposed haloes greater than 2.5 times the disk mass to stabilize cold disks. Because of the high halo mass, they did not find bar formation but got an impressive gaseous spiral response. This work will be discussed in depth in Chapter 4.

Over the years, one of the most productive users of grid codes has been Sellwood. He has used a polar grid code to study a wide variety of problems. As his work will be discussed extensively throughout the text, it will only be mentioned here. His most recent work (Sellwood, 1990) uses grid codes to explore groove instabilities that give rise to large-scale spiral structure.

Although not of direct influence on the present work, there is another important numerical approach—the nonlinear stellar orbit models of Contopoulos and Grosbøl (1986, 1988). These emphasize the role of periodic orbits, particularly the resonant orbits, in defining the basic characteristics of galaxies. They start with an observed rotation curve, which is used to derive the axisymmetric background potential. They impose a spiral perturbation potential on this background and calculate the stable orbit families in this potential. Orbits are given a Gaussian distribution of velocities about these central orbits and then recalculated. When making up the final response density, the orbits are weighted assuming an exponential disk. Two tests of self-consistency are

that the imposed and the response density maximum (per annulus) be the same, and that the amplitude of the imposed and the response two theta components match. For strong grand design galaxies, they find that the spiral ends at the 4/1 resonance. For weaker, tighter Sa galaxies, the spirals end at corotation. Work with a bar plus spiral perturbation is currently in progress.

The present work introduces a code to model both the stellar and the gas disk components. The gravitational forces are calculated on a Cartesian grid, while the gas forces are calculated utilizing a smooth particle scheme. Prior to this work, the code used to model galaxies at the University of Florida was based on the beam scheme (Sanders and Prendergast, 1974; Sanders and Huntley, 1976), a code with a number of limitations. In particular, the beam scheme was only used to model a massless gas component, and the viscosity inherent to the algorithm could not be easily controlled. In addition, the method used to excite spiral structure was an imposed oval distortion, a physically unrealistic device. All of these limitations have been overcome in the current code. Both components can be modeled (the gas can make up any desired fraction of the total mass) with the self-gravitational forces of the gas included in a self-consistent manner. The viscosity is easily adjusted and controlled through the use of two input parameters.

In addition to including both a gas and a stellar component, the mechanism used to excite bar and spiral structure is quite different from other model approaches. No imposed driving, or perturbing, forces are used. Instead, a method of stabilizing disks through the inclusion of velocity dispersions was developed. These dispersions are calculated

by taking velocity moments of the collisionless Boltzmann equation. Featureless, stable disks can be constructed by adding peculiar motion while decreasing the circular velocity. Bar and spiral arm formation can then be induced and controlled by reducing the stabilizing stellar velocity dispersions in the center of the model region, while maintaining them at the outer edge. A bar forms naturally in the interior, and spiral structure is excited in the outer gas disk in response to the bar. Realistic spiral structure is formed and maintained in the gas component for more than 12 rotation periods. Thus, no destabilizing or perturbing forces are imposed: rather a bar evolves naturally, in a self-consistent manner from a destabilized disk. By varying the size of the reduced heating region and the halo mass, a wide variety of bar and spiral structures can be created.

Toomre Disks

To start the calculations, particles can be distributed in a number of possible disks; usually a Kalnajs/Hohl disk or one of the family of Toomre infinite disks, most often n = 0. The Hohl disk has many analytical features that make it a useful test case but exhibits solid body rotation, thus limiting its value as a physically realistic galactic model. The Toomre disks have both astrophysical and numerically favorable qualities and most of the models presented in this work are initialized as Toomre n = 0 disks. The Toomre disks are a series of flattened, infinite, self-gravitating disks, developed by Toomre (1963). The n = 0 disk, also known as the generalized Mestel disk, is the first in the sequence. It is the only Toomre disk to exhibit a flat rotation curve, a feature found in many observed spiral galaxies. In the simulations, two versions of this disk are used—the infinite disk and the exact truncated disk. The bulk of the models are initialized as infinite disks,

which are simply truncated at a radius R. The surface density of these disks is given by:

$$\sigma_o(b,r) = \frac{C_1^2}{2\pi G\sqrt{r^2 + b^2}} \qquad (1.1)$$

where b is the shape parameter of the disk and C_1 is the asymptotic value of the rotational velocity, approached as the radius increases to infinity. The disk mass within radius r can be derived from equation (1.1) and is given by:

$$m_o(b,r) = \frac{C_1^2 r}{G}\left[\left(1 + \frac{b^2}{r^2}\right)^{\frac{1}{2}} - \frac{b}{r}\right]. \qquad (1.2)$$

The potential for the disk is:

$$\phi_o(b,r) = C_1^2 \ln\left[b + \left(b^2 + r^2\right)^{\frac{1}{2}}\right]. \qquad (1.3)$$

and the rotational velocity at r:

$$V_o^2(b,r) = C_1^2\left[1 - \frac{b}{\sqrt{r^2 + b^2}}\right] \qquad (1.4)$$

which approaches the asymptotic value of C_1 as r increases. On the left in figure 1-1 the surface density for two values of b, normalized to the $b = 4$ central value are shown. On the right the rotational velocity for the same two values of b are plotted. As b increases,
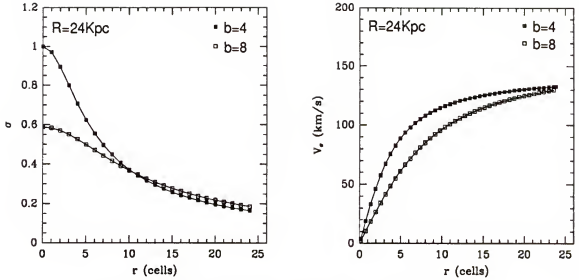
Figure 1-1: Surface density and circular velocity of an infinite
Toomre n = 0 disk for two values of the shape parameter, b.

the surface density falloff is more gradual and the rotational velocity rises less steeply.

When the n = 0 disks are used, it is necessary to truncate them at some radius R, the total disk radius. Their masses diverge, approaching ∞ as $R \to \infty$. Hunter, Ball and Gottesman (1984) derived exact solutions for the surface density and rotational velocities for truncated n = 0 disks. For such exact truncated disks, the velocity within the disk does not change from that given in equation (1.4) but the mass and surface densities have more complicated forms. The total mass interior to radius r is given by

$$m_o(b,r) = \frac{2C_1^2}{\pi G}[R - \sqrt{R^2 - r^2} - b\tan^{-1}\left(\frac{R}{b}\right) + \sqrt{r^2 + b^2}\tan^{-1}\left(\frac{\sqrt{R^2 - r^2}}{\sqrt{r^2 + b^2}}\right)],$$

(1.5)

where R is the truncation radius. The total disk mass is then:

$$M_o(b,R) = \frac{2C_1^2}{\pi G}\left[R - b\tan^{-1}\left(\frac{R}{b}\right)\right].$$

(1.6)

The surface density can be calculated from the mass to give:

$$\sigma_o(b, r) = \frac{C_1^2}{\pi^2 G} \left(\frac{1}{\sqrt{r^2 + b^2}} \right) \tan^{-1} \left(\frac{\sqrt{R^2 - r^2}}{\sqrt{r^2 + b^2}} \right). \tag{1.7}$$

Finally, the rotational velocity beyond the edge of the disk ($r \geq R$) can be calculated via:

$$V_o^2(b, r) = \frac{2C_1^2}{\pi} \left[\sin^{-1} \left(\frac{R}{r} \right) - \frac{b}{\sqrt{r^2 + b^2}} \tan^{-1} \left(\frac{R\sqrt{r^2 + b^2}}{b\sqrt{r^2 - R^2}} \right) \right]. \tag{1.8}$$

For $r >> R$, the rotational velocity approaches the Keplerian value

$$V_o^2(b, r) \approx \frac{2C_1^2}{\pi r} \left[R - b \tan^{-1} \left(\frac{R}{b} \right) \right] = \frac{GM_o(b, R)}{r}. \tag{1.9}$$

All truncated Toomre disks with indexes greater than 0 can be generated by successive differentiations, with respect to b^2, of the n = 0 result. Thus the rotational velocity, the mass, and the surface density are readily calculated for the n = 1 disk from the equations above.

The only other Toomre disk used to initialize calculations was the infinite n = 1 disk. Although this disk falls off in density more rapidly than the n = 0 disk, and thus might appear at first to be a better approximation of a truncated disk, it turns out not to be as useful for numerical reasons. The density drops so quickly in the n = 1 case that for a fixed number of particles the initial central densities are quite high, with approximately 250 particles in one of the central cells (compared to 40 for the n = 0 case). Such high cell densities can lead to numerical problems inherent to the grid scheme.

In the numerical algorithm, particles are laid out randomly in annuli, according to equation (1.2). A plot of a Toomre n = 0 disk, the initial configuration for most of the

models presented in this dissertation, is seen in figure 1-2. Rotation periods are given in terms of the cold rotation rate of this initial disk, at a point halfway across the disk.

The remainder of the thesis will present, in detail, the topics mentioned below. In Chapter 2, the numerical algorithm will be explained and tests of its soundness presented. Chapter 3 discusses techniques to stabilize a cold disk, while Chapter 4 introduces the methods used to excite bar and spiral structure. There is a change of focus in Chapter 5, which deals with the second part of the thesis, the gathering and reduction of 21cm HI radio observations of two barred spiral galaxies, NGC 1398 and NGC 1784. The observations were made with the Very Large Array of the National Radio Astronomy Observatory (NRAO). Lower resolution observations, using the C/D configuration, were gathered in 2/91. Higher resolution observations, made with the B/C array, were obtained in 2/92. The data sets were calibrated independently and then combined to make the final channel maps, H I density plots, and radial velocity contour maps. These observations, along with existing infrared surface photometry, will be used in the future to constrain the numerical models. Simulated observations, produced by the code, will be compared to the actual observations. Chapter 6 is used to discuss possible applications of the code and to summarize major conclusions presented throughout the dissertation.
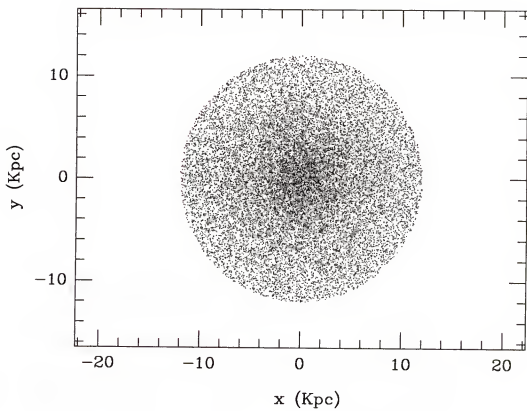
Figure 1-2: A Toomre n = 0 disk.

Outline of the Code

In this chapter the development and testing of a self-consistent, two-component computer algorithm to model spiral galactic disks is presented. The code is a two-dimensional hybrid grid/smooth particle scheme with the viscous forces for the gas component computed in a smooth particle manner, while the gravitational forces are calculated using a Fast Fourier Transform (FFT) on a Cartesian grid.

Grid codes have been employed to solve Poisson's equation since the late 1960s. Miller and Prendergast (1968), Hohl and Hockney (1969), and Hockney and Hohl (1969) were among the first to use them to model stellar galactic disks. Most of this early work was done on nondifferentially rotating Kalnajs disks with grid sizes up to 256×256 and 100,000 particles. Hohl was also the first to use grids to study differentially rotating disks. Miller (1976) extended the grid scheme to the more sophisticated polar grid codes. With these, he carried out extensive calculations, typically using up to 200,000 particles. By now, grid codes have been used to study a wide variety of disk models on a broad range of galactic and cosmological problems. Sellwood, who refined the polar grid technique, has used these grids to work on bar formation and evolution (1981, 1983), the effects of an imposed versus a self-consistent halo that interacts with the disk (1980), the effects of softening (1981), the effects of accretion on a stellar disk (Sellwood and Carlberg, 1984) and, most recently, groove studies (1990). A few investigators have attempted

three dimensional Cartesian codes (Hockney and Brownrigg, 1974; Miller and Smith, 1979). Due to the increased computer requirements of these codes, the highest resolution calculations of this type have been carried out on a 64^3 grid (Miller and Smith, 1979).

The primary advantage of a grid code over a direct N-body or smooth particle approach is speed. The number of computations necessary to calculate the force at all points on the grid is proportional to the number of cells in the grid, rather than the number of particles. For typical grids, this results in running times that are orders of magnitude less than either direct N-body calculations for a reasonable number of particles, or tree code simulations. As an example, a typical stellar calculation (20,000 particles) using a 64×64 grid code requires about an hour on a Sun Sparc2 workstation to complete 14 rotations. A similar calculation on a tree code runs for more than two days on the IBM 3090, a more powerful computer.

The price paid for this speed is not exorbitant for many cases of interest. The grid code has several limitations, none of which detract seriously from its usefulness in the study of the internal dynamics of galaxies presented here. First, the spatial resolution is limited by the grid size. Features smaller than a cell size are not guaranteed as real due to the technique used to calculate the gravitational potential. This procedure assumes that all material within a cell lies at the cell center. Thus, any distribution of matter within the cell is lost and the orbit of any individual particle is not followed reliably. A second effect of this grouping of material at the cell centers is that the results are most accurate in the limit of small potential gradients. The more constant the number density in cells surrounding a given cell, the better the estimate of the potential at that cell. In

addition, the accuracy of the potential at any cell, due to matter in any other cell, will be higher the farther away the other cell is, because the approximation of placing all matter at the cell center is less extreme the farther away that cell is. For these reasons, large density contrasts are not handled reliably, and a single grid should not be used to investigate questions such as the dynamics of interacting galaxies. Even a very strong bar surrounded by a significantly lower density disk may be problematic.

Both these limitations can be overcome through use of a larger (finer) grid. The latter problem can also be alleviated by a polar grid, which has smaller cell sizes in the center of the model where the density tends to be the highest. The major disadvantage of a polar grid is that considerably more effort is required to write the code. Another, less severe limitation of grid codes is that the grid size is fixed to be a power of two (or, with some extra coding effort, a number rich in factors of two, such as 96) due to the use of Fast Fourier Transforms. As the grid cannot be made any arbitrary size, control over the resolution is neither as easy nor as flexible as in a tree code. In addition, for a given model, the grid represents a fixed physical size, which cannot be adjusted once the calculation has begun. As particles move off the grid, they are lost from the calculation and cannot be tracked to arbitrarily large distances.

In a galaxy, it is believed that stellar motions are controlled by the overall, slowly varying gravitational force field of the galaxy as a whole rather than by the presence of other nearby stars. If this is true, the system is said to be collisionless. A simulated galaxy contains orders of magnitude fewer particles than an actual galaxy with each particle being many times more massive than a typical star. A consequence of this

is that the model particles suffer more frequent and more violent encounters than their counterparts in a real galaxy and the assumption of a collisionless system may be violated. The grid procedure of placing all the matter within a cell at the cell center helps to create a collisionless environment in which the potential for any particle is dominated by long-range forces rather than close encounters because when calculating the potential in a cell, the nearest surrounding matter is at least one cell away. The topic of collisionless systems will be returned to later in the chapter.

The viscous forces for the gaseous component of the disk are calculated in an entirely different manner. The gas is modeled as a continuous medium rather than as discrete particles, through the use of a smoothing (interpolating) kernel in a manner similar to that employed in smooth particle codes. Such codes were introduced in the 1970s (Gingold and Monaghan, 1977; Lucy, 1977) as Lagrangian methods which do not require the use of a grid, except as a computational tool. The method followed most closely in this code is that of Lucy, who applies Monte Carlo theory to calculate the continuous fluid density, velocity and their derivatives. Following Lucy, consider the problem of approximating a value for a function η at r, from a set of J discrete points randomly distributed within a volume V, i.e.;

$$\eta(r) = \int_V w(r - r')Q(r')\rho(r')dV', \qquad (2.1)$$

where w is the interpolating kernel or the weighting function. It acts over a small area of radius σ. Standard Monte Carlo theory states that if the set of J points, at r_j, is distributed such that the probability of finding a point in volume element dV' is

proportional to $\rho(r')dV'$, then

$$\tilde{\eta}(r) = \frac{1}{J} \sum_j w(r - r_j)Q(r_j) \qquad (2.2)$$

converges to $\eta(r)$ as the number of points, J, approaches infinity.

If it is required that w be normalized

$$\int_V w(r - r')dV' = 1 \qquad (2.3)$$

and that w is 0 for $|r - r'| > \sigma$, then $\eta(r) \rightarrow Q(r)\rho(r)$ as $\sigma \rightarrow 0$. It follows that

$$\tilde{\eta}(r) \rightarrow \eta(r) \rightarrow Q(r)\rho(r) \qquad as \quad J \rightarrow \infty$$

$$and \quad \sigma \rightarrow 0 \qquad (2.4)$$

The density and velocity at any point can be found by letting $Q = 1$ and $\vec{V}$ respectively.

The spatial derivative of the velocity is also needed in the viscous force calculation. The advantage of the smooth particle method is that Monte Carlo theory is applied to the discrete representation to obtain values for the continuous variables. The derivatives of the continuous variables can then be replaced by the derivatives of the smoothing function which can be obtained by analytical differentiation.

Thus the method used to calculate the viscous force is significantly different from that used to compute the gravitational forces. As it is necessary to search a small area of radius σ (in two dimensions) around each gas particle for all other gas particles, these calculations require substantially more computational time than do the gravitational force calculations. To run a model with 25,000 gas and 25,000 stellar particles and a halo through 14 rotations on a 128 × 128 grid requires approximately two days on a Sun Sparc2 workstation.

To summarize, both a stellar and a gaseous disk can be modeled in the current code. All particles feel a gravitational force, calculated using a Cartesian grid, while the gas particles feel an additional viscous force calculated in a smooth particle scheme. The gas makes up any fraction of the total mass and half the total number of particles. In addition to the disk components, a spherical halo can be included. The halo is imposed and therefore is not included in a self-consistent manner. It is used solely as a stabilizing agent.

The Gravitational Forces

To calculate the self-gravitational forces on the particles, the equation for the force exerted on each particle by all other particles must be solved for each particle pair:

$$F_{ij} = \frac{Gm^2(r_j - r_i)}{(\varepsilon^2 + |r_i - r_j|^2)^{\frac{3}{2}}},\tag{2.5}$$

where $m_i = m_j$ and ϵ is the softening length, discussed in more detail later in this chapter. At each timestep, the sum over $i = 1, 2....N_p$, where N_p is the total number of particles, of the forces to which the jth particle is subject must be computed. The most straightforward, but least efficient, way to do this calculation is by direct summation. To solve equation (2.5) for all particle pairs would require (N)(N-1)/2 calculations. As the number of particles increases the calculations become prohibitive, limiting N_p in most cases to less than 5000. For large N, the most expedient scheme for evaluating the forces is through use of the Fast Fourier Method, (FFM) which requires only on order $4N_g^2\left[1 + 2\log_2\left(4N_g^2\right)\right]$ calculations per time step where N_g is the number of cells per grid side used to perform the Fast Fourier Transform (FFT). Thus, the potential

calculation using the FFT method is independent of the number of particles. For a typical case which uses many fewer grid points than particles, the potential calculation takes less time than that required to advance the coordinates of the particles. This yields a computation time "proportional to N, with a fixed overhead for the potential calculation, an efficiency that can scarcely be improved upon" (Sellwood, 1987, p. 165). Because of this advantage in speed when using a large number of particles, the FFM is the technique employed in the present code.

To utilize the FFM, the model area is broken into a two-dimensional Cartesian grid ($N_x \times N_y$), with a total of K cells (K = $N_x \times N_y$). The technique can be used equally well on a polar grid, which has the important advantage that the cells are smaller in the center of the grid where the density tends to be the highest. Using the FFM, an estimate of the potential at the center of any cell is obtained by assuming that all particles within a cell lie at the center of that cell. Then the sum

$$\Phi_{ij} = \sum_{i=1}^{2N_x} \sum_{j=1}^{2N_y} C_{ij} M_{ij} \tag{2.6}$$

is evaluated, where

M_{ij} = the mass of the particles in the i,j cell.

C_{ij} = the convolution coefficients:

$$C_{ij} = \frac{G}{\sqrt{\varepsilon^2 + r_{i,j}^2}} = \frac{G}{\sqrt{\varepsilon^2 + (i - i')^2 + (j - j')^2}}. \tag{2.7}$$

The convolution coefficients depend only on the softening length, ϵ, and the positions of the cell centers, both of which are constant throughout a calculation. They need only be computed once, at the start of the program, and stored. The calculation is done over $2N_x$, $2N_y$ to prevent the inclusion of mass outside the grid into the potential calculation (aliasing). A basic assumption when using a FFT is that the function of interest (the mass distribution say) is a finite segment (one period) of an infinite periodic function, in both dimensions. The value of the potential from the transform will include the effect of the periodic mass distribution. To avoid this, the function must be isolated in space. This is done using a technique called "padding with zeroes", by using a grid four times as large as the area of interest. The mass is non zero only for i,j less than N_x, N_y.

$GC_{ij}M_{ij}$ is a good estimate of the negative of the potential at i,j due to material in another cell if the separation between the two cells is not too small. Discrete Fourier transforms provide an efficient method for evaluating the sums over cells in equation (2.6). The Fourier convolution theorem states that the Fourier transform of the convolution of two functions is equal to the product of their individual Fourier transforms. Thus,

$$\widehat{\Phi}_{ij} = (2K)^2 \widehat{C}_{ij}\widehat{M}_{ij} \tag{2.8}$$

where $\widehat{}$ represents a Fourier transform. In order to determine the potential, the Fourier transform of the convolution coefficients is calculated once and stored, and the Fourier transform of the mass distribution is calculated each time step. These are multiplied and back transformed to get the potential at the center of each cell, each time step.

The potential calculation can be done in this economical way due to the development in the mid-1960s of decimation-in-time Fast Fourier Transforms, from the work of J. W. Cooley and J. W. Tukey (1965). They popularized a method of computing the discrete Fourier transform of N points in $O(N\log_2 N)$ operations. Until this time, the standard calculations of Fourier transforms took $O(N^2)$. Following the notation of Press et al. (1989), the standard procedure for calculating the discrete Fourier transform of N points, i.e., solving:

$$F_k \equiv \sum_{j=0}^{N-1} f_j \exp(2\pi ijk/N) \tag{2.9}$$

where k ranges from 0 to N-1, was to rewrite this as:

$$F_k = \sum_{j=0}^{N-1} f_j W^{kj}, \tag{2.10}$$

where W, a complex constant, is defined to be:

$$W \equiv \exp(2\pi i/N). \tag{2.11}$$

The vector of f_j's was multiplied by the matrix W^{kj}, whose $(k,j)^{th}$ element is the constant W to the power $(k \times j)$, to produce a vector whose components are the F_k's. This matrix multiplication required N^2 complex multiplications. In 1942, Danielson and Lanczos showed that a discrete Fourier transform of length N can be rewritten as the sum of two discrete Fourier transforms, of length N/2; one formed from the even-numbered points of the original N, the other from the odd. Thus

$$F_k = F_k^e + W^k F_k^o, \tag{2.12}$$

where F_k^e is the kth component of the Fourier transform of length $N/2$ formed from the even components of the original f_j's, while F_k^o is the corresponding transform formed from the odd components. To calculate F_k^e will require $(N/2)^2$, calculations, F_k^o will require a further $(N/2)^2$ calculations and to multiply F_k^o by W^k requires N calculations. Thus the total number of calculations is $N + (N^2/2)$ versus N^2 for the original discrete Fourier transform. This property is then used recursively and F_k^e is written in terms of the transform of its $N/4$ even-numbered input data, F_k^{ee}, and odd-numbered data, F_k^{eo}, and so on.

If N is an integer power of 2, the Danielson-Lanczos Lemma can be continually applied until the data are subdivided down to transforms of length 1. There will be $\log_2 N$ subdivisions in all. Each one-point transform is just one of the input numbers f_n, for some n, i.e., $F_k^{eeoeooeoeoe\cdots} = f_n$ for some n. Which value of n corresponds to which pattern of e's and o's is determined by reversing the pattern of e's and o's, and letting $e = 0$ and $o = 1$. This gives, in binary, the value of n. To compute the FFT, the original vector of data f_j is rearranged in bit-reversed order, so that the individual numbers are in the order of the number obtained by bit-reversing j. These are the one-point transforms. Combining adjacent pairs gives two-point transforms. Combining the adjacent pairs of pairs gives the 4-point transforms, and so on, until the first and second halves of the whole data set are combined into the final transform. Each combination takes of order N operations, and there are $\log_2 N$ combinations, so the entire algorithm is of order $N\log_2 N$ for a one dimensional transform. For a two dimensional transform, on order $M\log_2 M$ calculations are required, where $M = N_x N_y$. The FFT used in the code

was provided by J. M. Huntley of Bell Laboratories.

Once the potential at the cell center has been calculated, a nine point cubic spline is used to determine the potential and force at any other position within the cell. The differencing of the potential to solve for the force smooths the field and degrades the resolution somewhat. To prevent this the FFT can be used to solve for the force directly but each force component must be determined separately (Sellwood, 1981). The particles are advanced using one of two methods; time-centered leap frog, (TCLF) or second order Runge-Kutta (RK). Both methods are second order, but the time centered leap frog integrator requires only one force calculation per time step, while the RK technique requires two. The TCLF method advances the particle positions once every time step and the velocities once a time step at the half time step (Miller and Prendergast, 1968).

The final choice between the integrators was in favor of the TCLF. This was the technique used initially but early tests on a marginally stable, featureless disk showed the RK models lost fewer stars off the grid than did identical TCLF models. This seemed to warrant further experimentation with the RK method, so an additional test was made of the disk response to an imposed potential. The results of this calculation, seen in figure 2-1, seemed to favor the RK method. Shown are the density distribution initially and after 10 rotations. By the latter time, the particles have moved from the original Toomre disk distribution into thin rings separated by one grid cell. As the imposed potential is calculated at the center of the cell, assuming that all the particles lie at the center of the cells, the particles migrate to the cell centers. The loss of angular momentum after 18 rotations is −0.30% for this model. In a similar test using TCLF, the final distribution
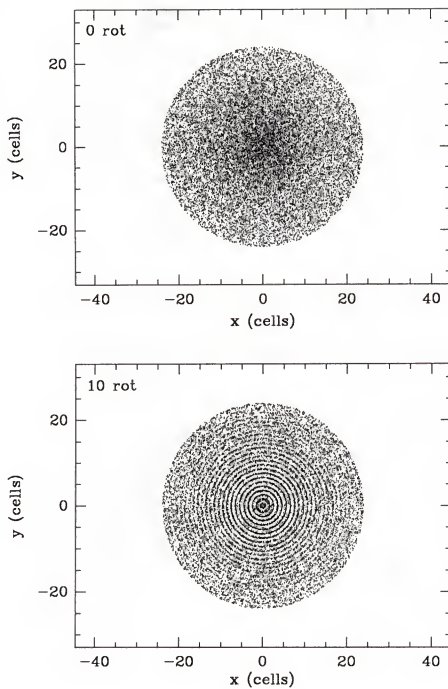
Figure 2-1: The disk response to a second order Runge-Kutte integrator.

differs very little from the initial and particles do not settle into such rings. However the change in angular momentum was only 3.2×10^{-4} %. It was not clear which integrator was doing the better job.

However, on featureless disk calculations that include the self-gravitational forces, the TCLF again better conserves angular momentum, but here the differences seem more significant. Identical models show a change in angular momentum of less than half a percent using the TCLF, but more than 5 percent using RK. In agreement with the initial tests, the RK model lost only 3.7% of its stars, while almost 6% of the particles moved off the grid for the TCLF. In addition, models using RK tend to have slightly higher densities in the center of the grid. It was later discovered, on disks with structure, that the conservation of angular momentum using the TCLF scheme was once more better than the RK case. In these cases the angular momentum losses could be significantly higher using RK. For models that included a gas component, it is not possible to calculate the gas viscous forces twice, as the RK technique requires and this may have contributed to the large angular momentum errors. In conclusion, for featureless stellar models, the differences between the two integrating methods did not seem significant, but when bar structure developed, the discrepancies, particularly in angular momentum conservation, increased and the decision was made to run the final, higher resolution models using TCLF.

The time step used to advance the potential is determined by the Courant condition, or a multiple of it. The Courant condition can be written as:

$$dt < \frac{L}{V_{max}}, \tag{2.13}$$

where L is the cell size in physical units. This assures that no particle travels more than one cell per time step, a condition required for numerical stability. The effect of varying this parameter is discussed below under variable numerical parameters.

If a particle moves off the grid, it is removed from the calculations permanently. A disadvantage of this is that these particles carry angular momentum with them. Despite this, no attempt is made to return lost particles back to the calculations, mainly because it is not clear how to do so in a physically realistic manner. All schemes that return particles to the computational area are in some sense contrived. As particles are not replaced, some thought must be given to choosing the initial radius to minimize particle loss. On one hand, it is advantageous to maximize resolution. On the other, it is expected in most of the models of interest, that the final configuration will be larger than the initial, and at least some particles will be lost.

The Gas Component

Although gas is a relatively minor component of spiral galaxies, typically less than five to ten percent of the total mass, it is of major importance in a computer simulation for two reasons. First, it is the more important component observationally. Many of the observational comparisons for a model come from HI 21 cm radial velocity profiles and density distributions. Secondly, because gas is dissipative, spiral structure is expected to show up much more strongly in the gas than in the stellar component. This is true for both actual galaxies and numerical models. For these reasons, the code includes both a stellar and a gaseous component. Both components feel gravitational forces but gas

particles differ from the stars through the application of a viscous force. The viscous force is a bulk viscosity added in a method following Sanders (1977). It contains just a single term, dependent on the square of the velocity divergence. It is applied to the gas flow only in the case of compression, thereby acting to prevent the crossing of orbits and allowing the particle trajectories to mimic gas streamlines. The viscous coefficient may be adjusted, rather than being implicit to the numerical algorithm (as is the case in the beam scheme).

The viscous forces are calculated in a smooth particle manner. The equation of motion for an individual "fluid" gas particle is:

$$m_i \frac{d\vec{V_i}}{dt} = -\vec{\nabla}(\Phi_i + q_i) \tag{2.14}$$

where Φ_i is the gravitational potential at the particle position and q_i is the viscous pressure. The viscous force acting on the jth particle is defined to be:

$$\vec{\nabla}q_j = \frac{1}{\rho_j} \sum_i^{n_\sigma} m \frac{q_i}{\rho_i} \vec{\nabla} w(|\vec{r}_j - \vec{r}_i|), \tag{2.15}$$

where

ρ_i, ρ_j are the continuous densities at the ith and j^{th} particle positions, $\rho_i = n_i m$, where n is the number/area.

m is the mass of the particle, all gas particles are assumed to have the same mass.

w is the weighting function.

q_i is the viscous pressure at particle i.

The summation is carried over all gas particles within a circular area of radius σ of particle j, n_σ. The viscous pressure introduced in the previous equation is specified as:

$$q_i = \alpha \sigma^2 \rho_i (\vec{\nabla} \cdot \vec{V})^2. \tag{2.16}$$

where α is a dimensionless number parameterizing the strength of the viscous force.

In order to calculate the viscous forces, the density and the velocity, as well as the velocity divergence, must be known at each particle's position. From equations (2.1) through (2.4), any macroscopic fluid variable, Q, at $\vec{r}_j$ can be written as

$$(\rho Q)_j = \sum_i^{n_\sigma} m Q_i w(|\vec{r}_j - \vec{r}_i|). \tag{2.17}$$

The density and velocity are found by setting Q equal to one and $\vec{V}$, respectively. Thus,

$$\rho_j = \sum_i^{n_\sigma} m w(|\vec{r}_j - \vec{r}_i|), \tag{2.18}$$

and

$$\left(\rho \vec{V}\right)_j = \sum_i^{n_\sigma} m \vec{V}_i w(|\vec{r}_j - \vec{r}_i|). \tag{2.19}$$

To determine the density at a particle's position, a summation is done over all gas particles lying within σ of the particle in question. Particles within this area contribute to the density according to their distance from the particle in question; i.e., they are multiplied by the weighting function, $w(\xi)$, where ξ is the separation between the two particles; $\xi = (\vec{r}_j - \vec{r}_i)$. The weighting function peaks at $\xi = 0$ and falls to zero at $\xi = \sigma$.

In order to treat the gas as a continuous fluid accurately, several other gas particles must lie within σ of any given particle. If this is true, the continuity of the derivatives

of the fluid quantities is guaranteed and can be defined in terms of $\vec{\nabla} w$ (Sanders, 1977). In general:

$$\vec{\nabla}(\rho Q)_j = \sum_i^{n_\sigma} m Q_i \vec{\nabla} w(|\vec{r}_j - \vec{r}_i|). \tag{2.20}$$

In particular, the velocity divergence can be written:

$$(\rho \vec{\nabla} \cdot \vec{V})_j = \sum_i^{n_\sigma} m(\vec{V}_i - \vec{V}_j) \cdot \vec{\nabla} w(|\vec{r}_j - \vec{r}_i|). \tag{2.21}$$

The viscous force is only calculated if $(\nabla \cdot \vec{V}) < 0$. Boundary conditions place restrictions on the form of the weighting function. To ensure continuous forces requires that (Sanders, 1977):

$$w(r \geq \sigma) = 0$$

$$\frac{dw(\sigma)}{dr} = 0 \tag{2.22}$$

$$\frac{dw(0)}{dr} = 0$$

In addition, normalization in two dimensions requires:

$$2\pi \int_0^\sigma w(\xi)\xi \cdot d\xi = 1 \tag{2.23}$$

One choice for $w(\xi)$ which satisfies these conditions is the third order polynomial, whose coefficients may be solved for using equations (2.22) and (2.23):

$$w(\xi) = \frac{10}{3\pi\sigma^2} - \frac{10}{\pi\sigma^4}\xi^2 + \frac{20}{3\pi\sigma^5}\xi^3. \tag{2.24}$$

A plot of $w(\xi)$ is shown in figure 2-2. Discussions of the effect of various weighting functions can be found in Schussler and Schmitt (1981) and Monaghan (1985).
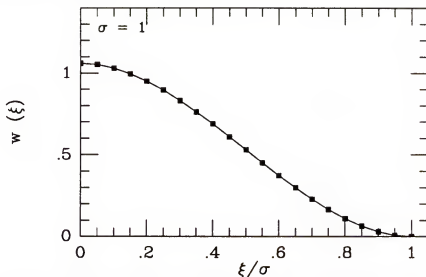
Figure 2-2: The weighting function as a function of particle separation.

To illustrate how the gas forces act as a dissipative mechanism, consider the simplest case of two gas particles, within σ of each other. If the particles are moving away from each other, no viscous force is applied. If, however, they are approaching each other head on, the viscous force will act in a direction opposite to the motion for both particles, tending to brake their motion. If the particles are travelling parallel to each other, with one overtaking the other, the forces will act in the direction of the motion for the slower moving particle and against the motion for the faster particle, again, preventing the "crossing" of orbits. In a real case, the particle motions are much more complicated than this, and there are many more particles involved in each calculation.

The effect of adding dissipative forces is shown in figure 2-3. A cold Toomre model is calculated twice, once as a stellar disk, and then as a gas disk. The gas parameters are $\alpha = 0.025$ and $\sigma = 0.6$ cells. A stabilizing halo of $M_H = 3\ M_D(R)$ is added, where R is the edge of the initial disk radius. The value of r_o is 0.2 R. Both models are calculated on a 128×128 grid with an initial radius of 36 cells. The surface density of the stars and gas are shown at 2 and 10 disk rotations. The stars are shown in the top panels, the gas below. Because the disk is cold initially, instabilities are expected to develop, causing spiral structure even with a stabilizing halo. In the stellar disk, the random noncircular motion rises rapidly, due to fluctuations in the potential caused by the perturbations, and all spiral structure is lost, smoothed out by large noncircular velocities within three rotation periods. In the gas component, however, dissipative collisions act to keep the gas cool, allowing spiral structure to remain as long as the model was run, 12 disk rotations.

This point is illustrated in figure 2-4, which compares the stellar and gaseous velocity dispersions for these two models. Stellar plots are shown on top, the corresponding gas figures directly below. On the left, the peculiar radial velocity squared is plotted against disk rotations for three radii; 9 (squares), 19 (circles), and 29 (triangles) cells. On the right, the peculiar radial velocity across the disk, after 10 rotations, is seen. It is apparent why spirals persist in the gas but not the stellar component. The stellar disk quickly heats up, with peculiar velocities rising from zero to a Q of approximately 1. (Recall from Chapter 1, that Q is the ratio of the radial peculiar velocity dispersion to the Toomre dispersion minimum.) The rotation curves for these models are flat across the outer disk with a value of approximately 365 km/s. In the gas component, even
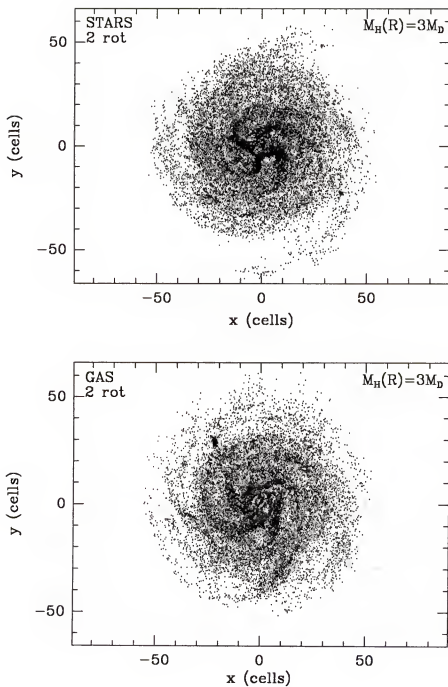
Figure 2-3: A cold stellar disk (top) and a cold gas disk (below).
A halo 3.5 times the disk mass is included in both models.
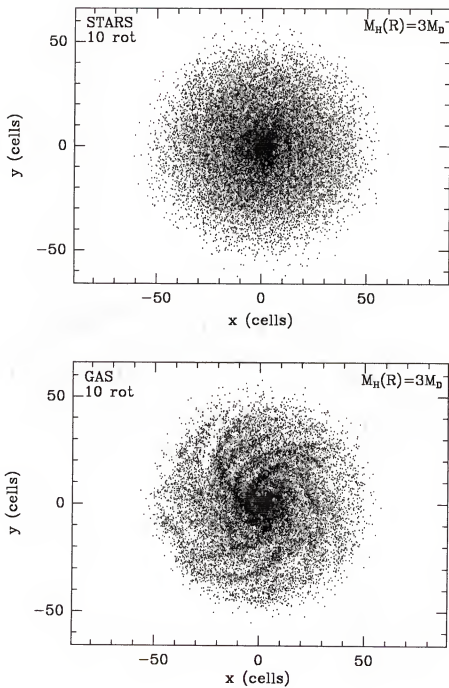
Figure 2-3: continued

for this low value of α, corresponding to viscous forces of only a few thousandths to a few hundredths of the gravitational and halo forces, the differences in the heating are significant, with the gas peculiar velocities at 10 rotations approximately half that of the stars due to viscous dissipation. It is the collisions between the gas clouds that act as a dissipative mechanism, preventing them from obtaining large random velocities. The low velocity dispersion allows the disk to be continually destabilized to the growth of spiral disturbances (Carlberg and Freeman, 1984) and robust spiral structure persists for many rotation periods.
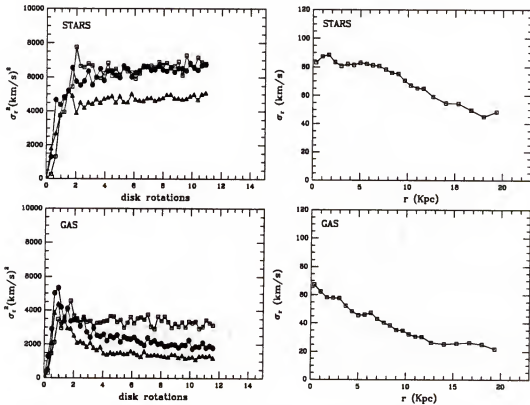
Figure 2-4: A comparison of stellar and gas velocity dispersions in cold disk models. Both models include a halo three times as massive as the disk.

Variable Numerical Parameters

In this section, an attempt is made to determine to what extent the code's results depend on numerical parameters such as the number of particles, the cell size, the grid size, the timestep, etc. When doing numerical experiments, it is important to know how results reflect the physics of the system being studied rather than some form of experimental error. The tests were done on a marginally stable, featureless, Toomre $n = 0$ disk. Compared are the variations in the number of stars lost off the edge of the grid, the maximum density per cell at the densest regions of the grid, the difference between the initial and final mass distributions and tests of angular momentum and energy conservation. Angular momentum and energy tests do account for particle loss, i.e., the angular momentum and energy of particles as they leave the grid is subtracted from the total initial value of these quantities. These numbers are compared to the final angular momentum and energy. The angular momentum and energy checks are obviously the only of these tests that can label one model as more accurate than another. The other tests are more subjective, but still useful in comparing one model to another. In addition, a parameter such as the number of particles lost depends on the initial radius compared to the total grid size, and it is meaningless to compare this for cases where the radius is a different fraction of the total grid size. The same is true for the maximum density per cell. These limitations aside, as these models are expected to be marginally stable, it was informative to compare the above quantities as useful tests of the stability of the code to varying numerical parameters. It was to be hoped that model results would not prove to be highly dependent on these numerical parameters, and to a large extent this was found.

Number of Particles

When modeling a galaxy, physically realistic results about a system on order of 10^{11} stars are sought using many fewer particles than this. How does this effect the results? Does the use of such small numbers have numerical effects? Does it have physical effects; i.e., the mass of each particle is much larger than a typical star but does this influence the results? The biggest concern about the number of particles is whether a system with so few particles is collisionless. The relaxation time decreases as the particle number decreases because the number of collisions and the importance of each increases with particle mass. For a two dimensional disk, the ratio of the relaxation time to the crossing time is proportional to the number of particles, the softening length and the ratio of the peculiar velocity to the circular velocity (Rybicki, 1972). Energy and angular momentum will only be conserved along a particle orbit if the system is collisionless. It is expected that the system will better approximate a collisionless system as N is increased. A second way to create a collisionless system is to add softening, as will be discussed later in this chapter.

For the gravitational forces calculated by the FFT, model results were found to be relatively insensitive to the number of particles, N. This is especially true for models that used a time centered leap frog integrating scheme. Tests were made using a 64 × 64 grid on a featureless, warm Toomre n = 0 disk with N ranging from 5000 to 80,000. Table 2-1 gives the percent of stars that have moved off the grid, and the changes in angular momentum and energy, after 12 rotations, for five values of N. The angular momentum conservation improves as N increases but even for N = 5000, the change is

less than 2%. The energy fluctuates a good bit more than the angular momentum for a given model, and there is no systematic trend found in the conservation of energy as N increases. The angular momentum and energy tests were found to be slightly machine dependent and will vary from one computer to another. However, the general trend of better conservation of angular momentum and no trend in energy conservation as N increases was found on all computers.

Table 2-1: Star loss and changes in angular momentum for increasing values of N.

N	stars lost (%)	change in a.m. (%)	change in energy (%)
5000	8.9	-1.78	0.35
10000	7.6	-1.16	-1.68
20000	5.3	-0.48	-3.71
40000	7.8	-0.97	-2.55
80000	6.5	-0.33	-3.08

[a] After 12 rotations.

Figure 2-5 shows the plot of mass fraction versus radius for the initial distribution, and after 12 rotations for various values of N. There is little variation in the final mass fraction curves once N increases to 10,000. Figure 2-6 shows the rotation curves after 12 rotations for four values of N; there is little difference between the results with the exception that the statistical fluctuations are smaller for larger N. The rotation curves are calculated by breaking the disk into rings approximately one cell wide (for a 64 × 64 grid, 2 cells are used if the models are calculated on a larger grid) and averaging the tangential velocity within the ring. As N increases the statistics improve and the rotation curves for the lower values of N will not be as accurate. Plotted but not shown were the

peculiar radial velocity squared versus disk rotation for all values of N. No systematic variations were found although statistical fluctuations again decrease as N increases.
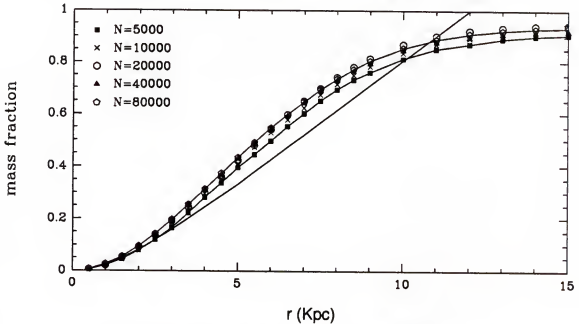
Figure 2-5: Mass fraction versus r for five values of N. The curve with no symbols is the initial distribution. Data points are connected for N=5000 and N=80000, the lowest and highest N values.

An identical set of models were calculated using a 2nd order RK integrator. The percent of particles lost was lower, for all values of N, than the TCLF models (by less than 3% for all cases), but the changes in angular momentum larger. In addition, deviations among models as N changed were slightly larger than for the TCLF case and the angular momentum conservation did not improve systematically as N increased.

The results presented above indicate that the models are relatively insensitive to N; still, there are a number reasons to have as many particles as possible. The model is more physically realistic as the number of particles increases, the relaxation time increases, the
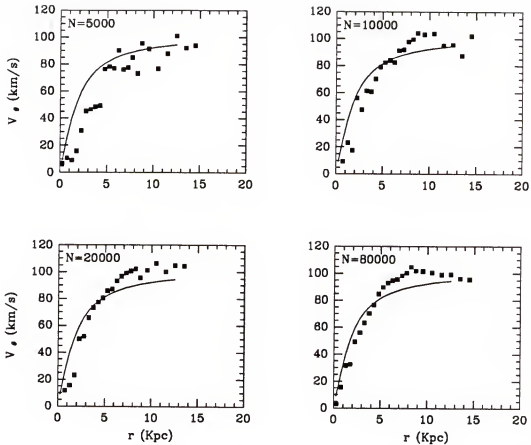
Figure 2-6: Initial (solid line), and final (filled squares), rotation curves for various values of N.

orbits should be followed more accurately and the statistics improve. The system gets closer to a collisionless system as N increases, however, as the numbers used are on order of 10^6 lower than the number of stars in an actual galaxy, a change by a factor of two, say, seems to be unimportant if N is large. As many of the useful tests done on a model include some type of statistical calculations in annuli across the disk, most of the models presented use N between 20000 to 40000 for the 64×64 grid cases and 40000 to 60000 for the 128×128 grid. These numbers are large enough to ensure sound statistics.

When calculating models which include a gas component, there is an extra concern about picking N. For the viscous forces to be calculated accurately, each gas particle must have several "near neighbors", the quantification of near being an input parameter. This sets a value for the minimum number of gas particles that can be used, which also sets the number of stellar particles. In addition, for the models in which the gas is laid out uniformly while the stars are in a Toomre disk distribution, some care must be taken to ensure that there are enough gas particles in each cell without increasing the stellar densities too much at the center of the model.

Cell Size

Each cell of the grid represents a physical length. This length, L, can be varied to allow for different sized systems. If the value L is changed without changing any other parameters, it simply acts as a scale factor and does not affect the outcome of the model. For example, for a Toomre disk, if the radius, R, of a disk is 24 cells, and L = 500 pc, the physical size of the system is 12 Kpc. The results from this model will be identical to a model with R = 24 cells and L = 1000 pc except that the velocities, forces, etc., will be scaled up because it is a physically smaller system having the same mass. The mass fraction across the disk, the ratio of the velocity dispersion squared in the radial and tangential directions in an annulus, the number of stars lost, the change in angular momentum, the position of the center of mass, etc., will be identical (within numerical precision) in both models.

Grid Size

Probably the most important parameter of a grid code is the number of grid cells. Due to the nature of the grid scheme, the resolution of a model is limited by the grid size to be no finer than one grid cell and features smaller than a few grid spacings are not to be trusted. In order to get higher resolution, or more detail in a model, a larger grid must be used. The use of a grid limits the resolution in two ways. First, when calculating the potential, all the matter in a cell is assumed to be located at the center of the cell and therefore detail about the distribution of matter within the cell is lost. The larger the total grid size, and thus the smaller percent of the total radius one cell represents, the better the resolution. Secondly, the FFT provides a good estimate of the potential at a cell center due to matter in another cell only if the two cells are not proximate. As the total number of cells increases as the grid size increases, the percent of cells for which the potential estimation is a good one also rises, yielding more accurate results. Unfortunately, with higher resolution comes an increase in both the computational time and memory required to run a job. Because the grid code uses a FFT to estimate the potential, the grid size is limited to be a power of two, or a number rich in factors of two, so the next grid size above 128×128 is 256×256. Further increases in running time occur as the grid gets larger because the number of time steps required to complete a set number of rotations rises and the number of particles must increase in order to maintain a reasonable number density. These factors dictated that with the current computer resources, the maximum grid used on a routine basis be 128×128. If only the stellar disk is modeled the 256 grid is possible (but slow), but if the gas component is included these models become

prohibitive, both in terms of computer time and storage. No gas models, and only a handful of stellar models, were run on this large grid.

Figure 2-7 shows the surface density of a stable Toomre n = 0 disk for a 64 × 64 and 128 × 128 grid. The lower resolution grid is on top. Models are shown after 12 rotations. To keep the number density per cell constant, N is increased from 10,000 to 40,000. For plotting purposes, only half of the 128 grid particles are included. Figure 2-8 shows the same disk run on a 256 × 256 grid, using 160,000 particles, at 12 rotations. Only a quarter of the particles are plotted. There is little variation among the figures. Plots of the peculiar velocity squared versus disk rotations show almost no differences for the three grids. The final mass fractions and rotation curves vary slightly between the grids but not in a systematic fashion. Table 2-2 gives the changes in angular momentum and the number of particles that have moved off the grid for these models. The angular momentum conservation gets consistently better as the grid size increases, however, these results indicate that for the featureless models, the 64 × 64 grid is a reasonable compromise between resolution and convenience.

Table 2-2: Star loss and the change in angular momentum, after 10 rotations, for three grid sizes.

Grid	stars lost (%)	change in a.m. (%)
64 × 64	3.63	-0.138
128 × 128	4.82	-0.097
256 × 256	5.20	-0.028
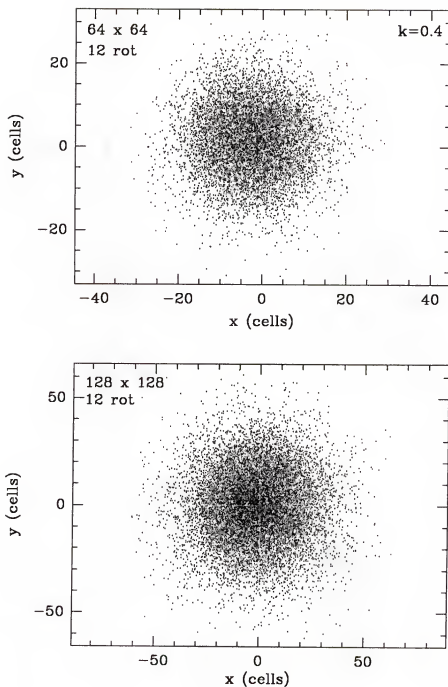
Figure 2-7: A stable stellar disk calculated on a 64 × 64 grid (top), and a 128 × 128 grid (below).
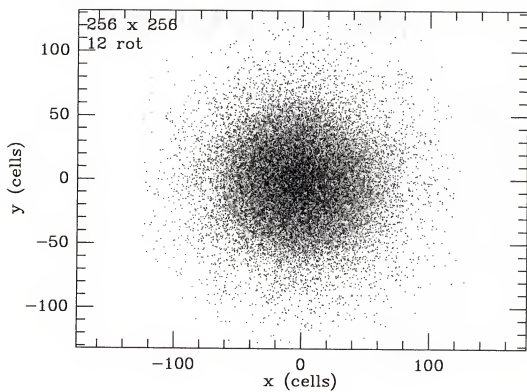
Figure 2-8: A stable stellar disk calculated on a 256 × 256 grid.

In the case of a stable featureless disk, the increase in grid size did not affect the results significantly. This is not surprising as there is little detail in the models. The grid size is expected to be a far more critical factor in models which form bar and spiral features. However, such models include both a stellar and a gas component. When gas is added there is a tremendous increase in both time and storage required to compute a model, limiting these models to grids of 128 × 128 or smaller.

Shown in figures 2-9 through 2-12 are two models expected to form bar and spiral features in the gas component, run on both a 64 × 64 and a 128 × 128 grid. For the smaller grid, 10000 particles are used, for the larger, 40000.

The first model is of a cold disk with a halo 3 times the disk mass within the initial disk radius included. The 64 × 64 grid results are shown in figure 2-9, the 128 × 128 in 2-10. In figure 2-9, the plots are on a smaller scale to increase the number density to that of figure 2-10. The second model (figures 2-11 and 2-12) is a cold disk with a halo of equal mass to the disk at the edge of the initial radius.

For both these models, the results are qualitatively the same for both grids. In both, the stellar component loses its spiral structure fairly quickly. The bar forms at about the same time and is about the same size. However, the detail in the gas component is much sharper in the higher resolution models. As a result of these tests, all of the two component models expected to develop structure were run using a 128 × 128 grid.

A final class of models, perhaps of only academic interest, are completely cold stellar Hohl disks. Such disks are known to be unstable to arbitrarily short wavelength perturbations and the instability develops more rapidly as the wavelength decreases. Thus

Figure 2-9: A two-component model with a halo three times the disk mass calculated on a 64 × 64 grid.
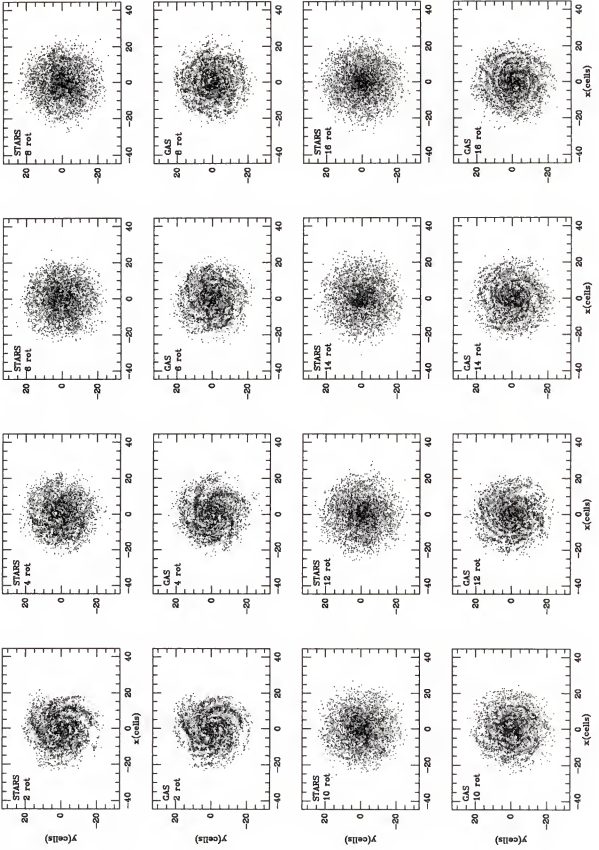
Figure 2-10: A two-component model with a halo three times the disk mass calculated on a 128 × 128 grid.
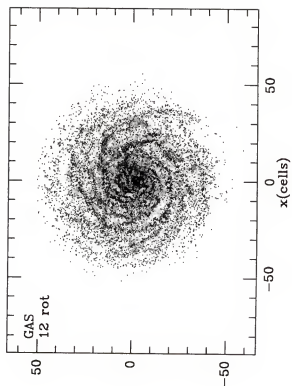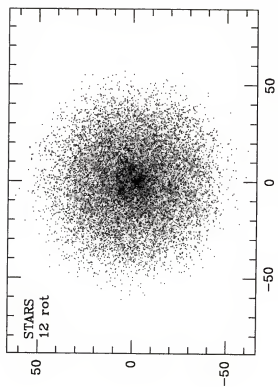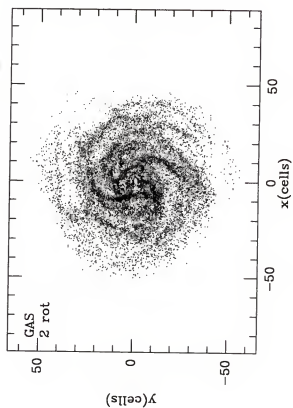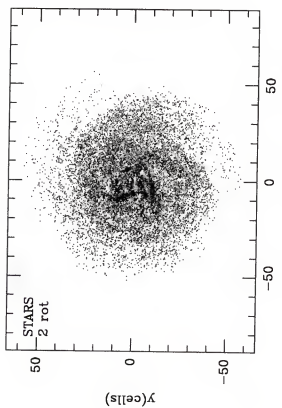
Figure 2-11: A cold, two-component model with a halo mass equal to the disk mass calculated on a 64 × 64 grid.

STARS 2 rot.
STARS 4 rot.
STARS 6 rot.
STARS 8 rot.
GAS 2 rot.
GAS 4 rot.
GAS 6 rot.
GAS 8 rot.
STARS 10 rot.
STARS 12 rot.
STARS 14 rot.
STARS 16 rot.
GAS 10 rot.
GAS 12 rot.
GAS 14 rot.
GAS 16 rot.
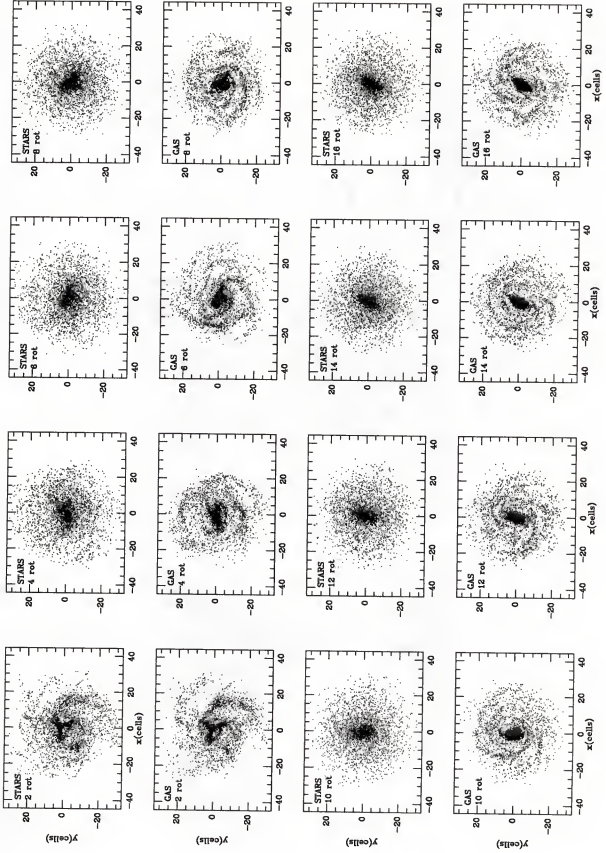
Figure 2-12: A cold, two-component model with a halo mass equal to the disk mass calculated on a 128 × 128 grid.
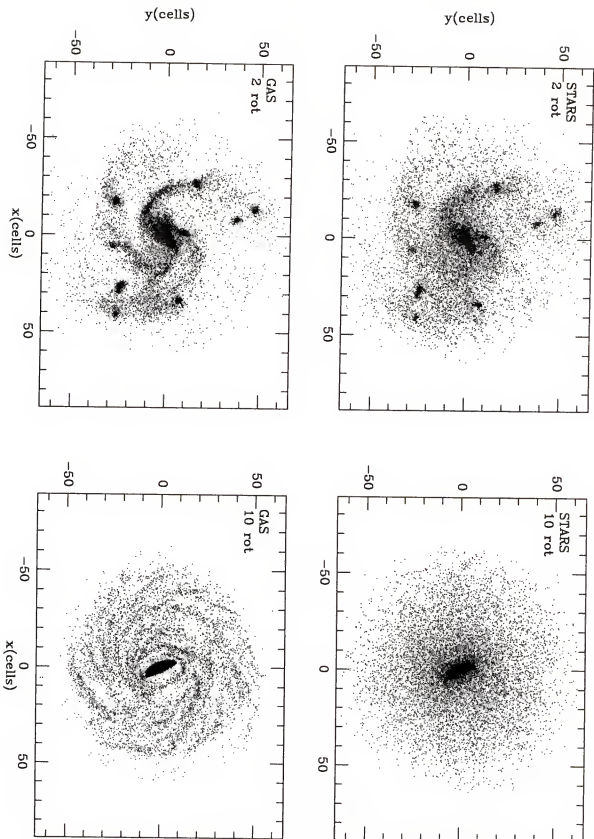
they provide excellent illustrations of the change in resolution that occurs when the grid size is varied. A cold Hohl disk was calculated for the three grids. The higher resolution grids allow smaller instabilities to develop as the smallest condensations possible are of the dimensions of a cell size. A courser grid eliminates the smaller condensations and in this sense, the grid adds "heating" to the disk. This is seen in figures 2-13 (64 × 64), 2-14 (128 × 128) and 2-15 (256 × 256). With the courser grid (64 × 64), the shortest wavelength perturbations are smoothed out and the very small-scale structure, caused by these perturbations, and seen in the higher resolution results, are absent. In addition, it takes longer for structure to clearly emerge. The condensations are much finer and develop earlier on the 128 × 128 and 256 × 256 grids. It is known that the short wavelength modes can be stabilized by the effects of heating, and use of a courser grid act like such a stabilizing velocity dispersion in this case.

Softening Length

In equation (2.5), the force is not simply the Newtonian gravitational force, but a "softened" force. Softening is used to make the system collisionless and to suppress the heating effects of close encounters. Using a softened force, the separation between two particles is $\sqrt{r^2 + \epsilon^2}$, rather than r. This greatly modifies the force at short distances. There are both theoretical and numerical reasons to add softening, so ϵ is not strictly a numerical parameter. Theoretically, in a galactic disk, the motion of a particle is expected to be governed by the long range effects of the more distant particles rather than by nearby encounters, i.e., it is a collisionless system. Because a numerical model is forced to use many fewer particles than an actual galaxy has stars, it is not as good an

Figure 2-13: A cold Hohl disk calculated on a 64 × 64 grid.
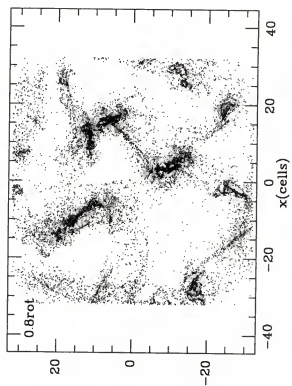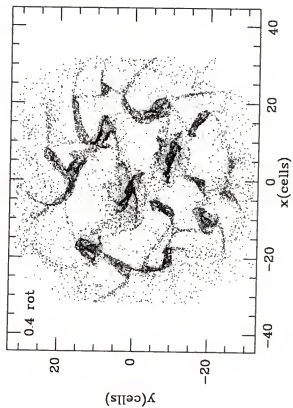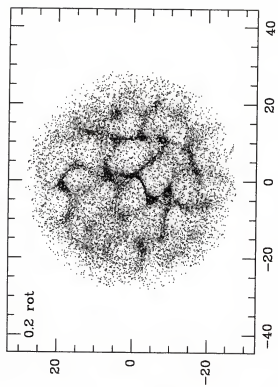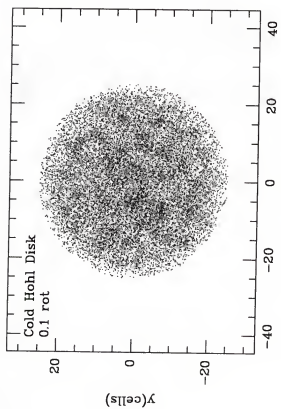
Figure 2-14: A cold Hohl disk calculated on a 128 × 128 grid.

Cold Hohl Disk
0.1 rot

0.2 rot

0.4 rot

0.8 rot

128 × 128

x(cells)

y(cells)
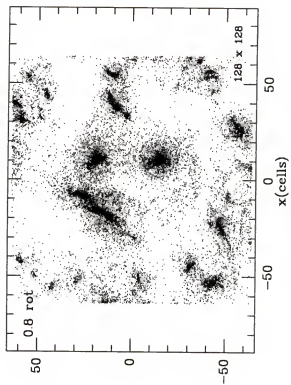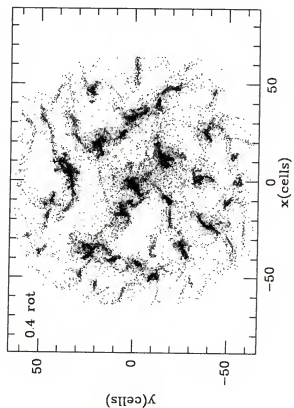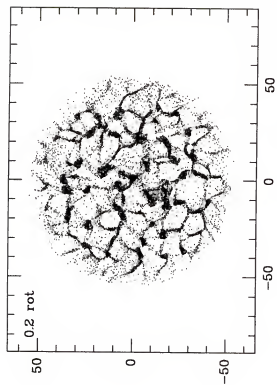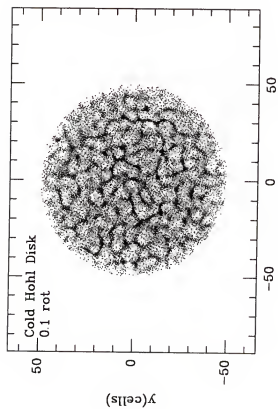
Figure 2-15: A cold Hohl disk calculated on a 256 × 256 grid.

64

Cold Hohl Disk
0.1 rot

0.2 rot

0.4 rot

0.8 rot

256 x 256
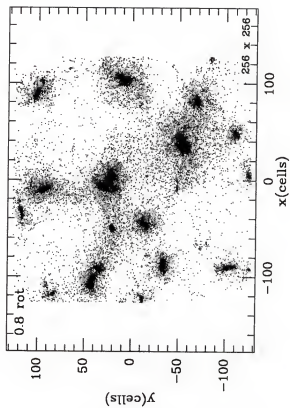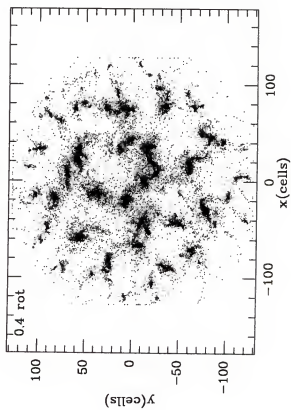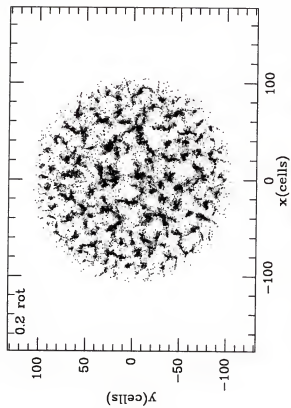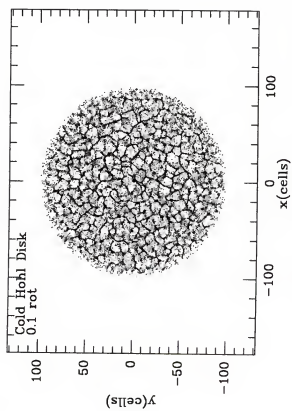
approximation to a collisionless system. Softening can be used to "suppress the small scale, steep fluctuations that characterize the field of a set of point masses and that are largely responsible for relaxation" (Sellwood, 1987, p. 157). As the softening length decreases, closer encounters are allowed, which increase the forces significantly, thus decreasing the size of the time steps and increasing the time required to run a model.

Due to nature of the potential calculations, softening is implicit in the grid approach and any explicit softening, represented by ϵ, is in addition to this. A grid code is therefore not as sensitive to the softening parameter as an N-body simulation or a tree code would be. When calculating the potential, it is assumed that all the matter within a cell lies at the center of the cell. Thus, when the effect of matter outside of a particular cell is being calculated, the nearest matter is seen to be in the center of a surrounding cell, i.e., at least one cell away.

Sellwood (1981) finds, using a polar grid code, that increasing the softening length has the effect of stabilizing models initialized with relatively small amounts of peculiar velocities ($Q \approx 1$) against bar formation. The large softening lengths worked against the gravitational collapse into a bar. A set of models expected to form a weak bar was calculated with an initial radius of 24 cells, allowing the softening length to vary from 0 to 5 cells. As the softening length increases, model results seem to improve; there is in general better energy and angular momentum conservation, fewer stars lost (see table 2-3) and, as seen in figure 2-16, less tendency toward bar formation, due to the weakened gravitational forces. However, as the softening length increases to very large values, these trends are reversed. Energy and angular momentum conservationdecrease

Figure 2-16: The effect of softening on a marginally unstable disk.
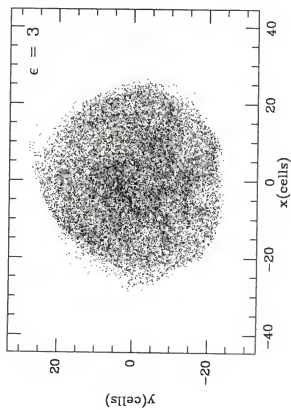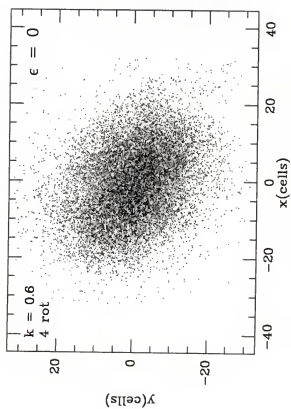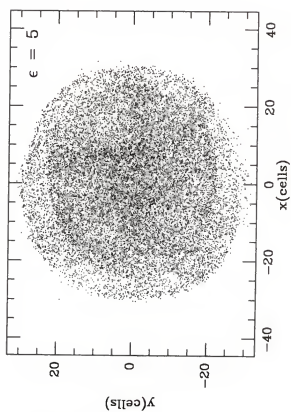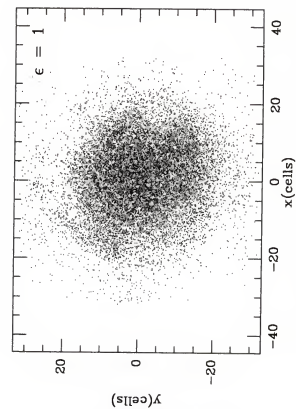
and many stars are lost off the grid at later times. Table 2-3 gives the number of stars

lost, the change in angular momentum and energy at 20 rotations, and the number of

time steps necessary to reach 20 rotations.

Table 2-3: Star loss, and the changes in angular momentum and energy at
20 rotations for increasing values of the softening length. The far right column
shows the number of time-steps necessary for the model to reach 20 rotations.

ϵ (cells)	stars lost	del a.m.	del energy	timesteps
0.0	13.02	-2.55	-2.07	3446
0.5	12.11	-4.28	-1.44	3394
1.0	10.93	-1.01	0.27	3195
1.5	8.46	-3.12	1.28	2989
2.0	6.87	-0.64	3.09	2794
3.0	3.22	-0.63	5.06	2455
4.0	3.67	-0.64	10.26	2232
5.0	30.83	-11.8	42.5	2021

Figure 2-17 shows the peculiar radial velocity squared versus time for the various

amounts of softening. The effect of the softening inhibiting heating by close encounters

is clearly seen. As the softening parameter increases, the peculiar velocities drop

systematically at all radii. Values are shown at 4.5 (squares), 9.5 (circles) and 14.5

(triangles) cells, for a model with an initial radius of 24 cells. A further example of

this is found in figure 2-18 in which a model identical to that presented in figure 2-3

but with an ϵ of 3 cells is shown. In figure 2-3, ϵ was 0.5 cells. Both models were

calculated on a 128 $\times$ 128 grid, with an initial radius of 36 cells. In 2-18, stellar spiral

structure is quite strong and obvious as 2 rotations, and although weak, is still apparent

at 10 rotations. The increased value of ϵ kept the velocity dispersions small enough to

prevent the destruction of the spiral structure by stellar heating. Finally, the effect of softening on the time step is evident in the last column of table 2-3.

Despite the apparent "improvement" in results that appear as the softening length is increased moderately, the softening length was seldom made larger than 0.5 cells on a 128×128 grid. For physical reasons, the softening length should be a few times the average interparticle length. For most of the two-component models presented in this work, ϵ was kept small, although tests were run with it up to 2 cells.
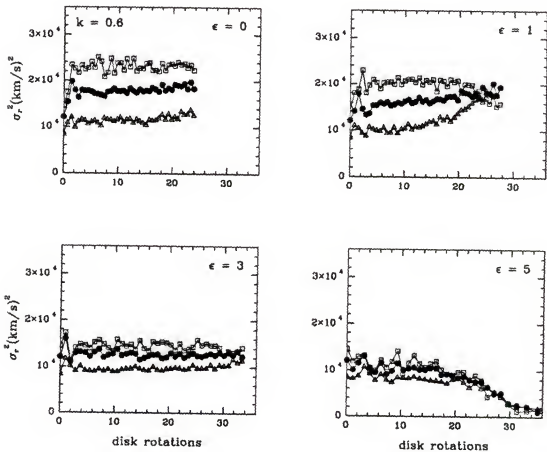
Figure 2-17 The peculiar radial velocity squared versus
disk rotations for four values of the softening parameter.
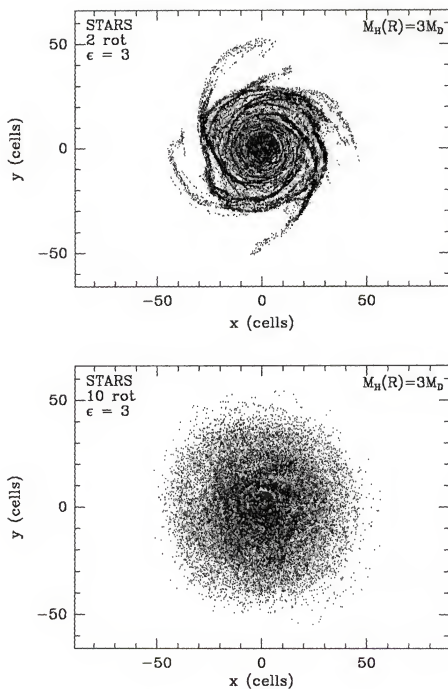
Figure 2-18: The effects of softening on a cold stellar disk with a halo 3 times the disk mass.

Time Step

The final parameter tested was the time step. The basic time unit is determined by the Courant condition, and the dimensionless time step, fcrnt, can be any multiple of this unit. If the dimensionless time step is 2, then the maximum distance any particle can move in one time step is two cells, etc. Tests were calculated changing this parameter from 0.5 to 8.

Table 2-4 shows the numbers of stars lost off the grid and the change in angular momentum and energy after 24 rotations as the time step increases. Until the value increases above 3, no systematic trends in these parameters are found. What seems most interesting is that even when grossly violating the Courant condition, the models do not deteriorate too much. Density plots after 24 rotations look similar for all the models. This stable behavior is probably due to the fact that these are featureless models, and

Table 2-4: Star loss and the change in angular momentum
after 24 rotations for increasing values of the time step.

fcrnt	stars lost (%)	del a.m. (%)	del energy (%)
0.5	8.9	-0.33	-0.70
1.0	7.8	-0.99	-1.45
1.5	8.1	-1.32	-0.87
2.0	9.0	-2.89	0.52
2.5	10.0	-2.70	1.74
3.0	7.4	-0.57	0.17
4.0	8.1	-0.79	1.8
5.0	9.8	-1.33	5.1
6.0	10.2	-2.12	8.7
8.0	15.7	-3.18	21.6

thus the potential varies smoothly from cell to cell. Tests run on bar models do suggest that it is better to use a smaller value of the time parameter and this parameter was seldom increased beyond 1-1.2.

Having documented the changes that occur from systematically varying parameters, it is interesting to ask what these numbers imply. How important is it if the angular momentum changes by half a percent between two models? To get an idea of the statistical fluctuations inherent in the code, a series of models were calculated changing the time step by a very small amount, in the second or third decimal place. The change in angular momentum is plotted in figure 2-18, for these results. As seen there is a healthy amount of statistical fluctuation.
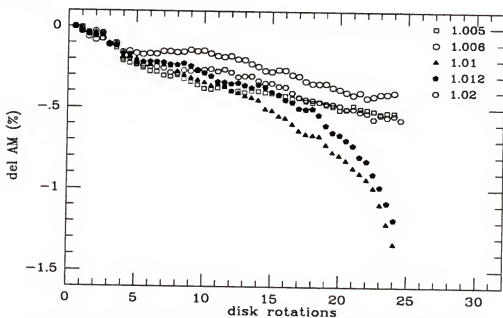
Figure 2-19: Statistical fluctuations in the change of angular momentum

As all the tests above indicate, model results are relatively insensitive to numerical parameters, as they should be. Although it does not in any way guarantee the correctness of a model, it is comforting that models do not vary much when a numerical parameter is changed. However, it is important to realize that these test were done on a small sample of disks, which may not have been sensitive test cases for all the parameters. Obviously, different disks may provide better tests of certain parameters. For example, the cold disks for the grid tests were much more sensitive than the warm disks. In numerical studies, it is better to err on the side of caution and to run a set of tests like the ones presented in this chapter on any promising models, before drawing any scientific conclusions from model results.

CHAPTER 3
STABILIZING AGENTS

Cold, flattened, self-gravitating disks, in which the stars move on strictly circular orbits, are violently unstable. Jeans (1919) showed that a homogeneous, three dimensional medium of infinite extent is unstable to gravitational collapse only if the perturbation wavelength is longer than the Jeans length;

$$\lambda_J^2 = \frac{\pi C_s^2}{G \rho_o}.$$ (3.1)

For a stellar system, the sound speed, C_s, is replaced by the radial velocity dispersion, σ_r. Flattening the disk to an infinitely thin sheet changes the problem to give a dispersion relation (Toomre, 1964):

$$\omega^2 = k^2 C_s^2 - 2\pi G \Sigma |k|,$$ (3.2)

where Σ is the constant surface density. The system is stable if $\omega^2 \geq 0$, i.e., if

$$\lambda < \lambda_J = \frac{C_s^2}{G \Sigma}.$$ (3.3)

As in three dimensions, if the sound speed or dispersion is not zero, the system is subject to gravitational instabilities at long wavelengths, but there is a fundamental difference between the three dimensional cloud and the flattened disk. For a completely cold, flattened system, in which the velocity dispersions (or sound speed) are zero, equation (3.2) reduces to $\omega^2 = -\frac{4\pi^2 G \Sigma}{\lambda}$. The cold disk is always unstable, and the instability grows more rapidly and more violent as the wavelength of the perturbation

74

decreases. Thus, the shortest wavelength perturbations are the most unstable, and a cold disk is unstable to all perturbations of arbitrarily short wavelengths. This result differs fundamentally from a cold three dimensional case, which is unstable, but for which the growth rate is independent of the wavelength of the perturbation.

When rotation is included, a term is added to the dispersion relation which works against the gravitational collapse, but this alone cannot stabilize cold disks. For solid body rotation, the dispersion relation becomes (Toomre, 1964):

$$\omega^2 = 4\Omega^2 - 2\pi G\Sigma|k| + k^2\sigma_r^2 \tag{3.4}$$

where the last term is zero for a cold disk. If differential rotation is present, this result can be generalized and written in terms of the epicyclic frequency, κ. The uniformly rotating result is a limiting case in which $\kappa = 2\Omega$. When compressed slightly over a small region of dimension L, the cold disk will collapse due to excess gravitational attraction if L is shorter than a critical length:

$$L_c \leq \frac{\pi^2 G\Sigma}{\Omega_{loc}^2} \tag{3.5}$$

(Toomre, 1964). Equation (3.4) shows that the disturbances of short dimension remain unstable despite rotation. The magnitude of L_c for a typical galactic disk is on order of the radius of the galaxy itself and perturbations of wavelengths shorter than this will destabilize the disk. Thus for flattened disks both pressure and rotation work against gravitational collapse. In a nonrotating sheet, pressure will work to stabilize short wavelength perturbations, while rotation decreases the size of the unstable regime by stabilizing the longer wavelength perturbation. If both are present in a disk, equation

(3.4) implies that the disk should be stable to all wavelength perturbations if $\frac{\sigma_r \Omega}{G\Sigma} \geq \frac{\pi}{2}$. This is from a linear treatment and gives a value of the velocity dispersion similar to the Toomre dispersion minimum, discussed in greater detail later in the chapter.

Numerical studies on a variety of cold flattened disks confirmed them to be unstable. Miller and Prendergast (1968) and Hohl and Hockney (1969) developed the initial grid codes and were among the first to apply numerical techniques in a realistic way to the study of cold stellar systems. Hohl and Hockney experimented with cold rigidly rotating (Kalnajs) disks which they found to be wildly unstable. Shown in figure 3-1 is a cold model of a Toomre n = 0 disk run on a 128 × 128 grid. It is obviously quite unstable, by one tenth of a rotation, particles have already started to clump. As the model progresses, the clustering gets stronger, and small filaments are present by 0.2 rotations. The pieces begin to coalesce into larger groups that are moving outward. The last plot shows the model at 0.8 rotations, and by this time there are six distinct clusters, all moving off the grid. At later times, a few of these pieces merge but continue to move slowly off the grid. By two rotations more than a third of the stars are lost from the calculation.

In 1971, Hohl carried out a major numerical study on Kalnajs disks, using a 100,000 body FFT code on a 256 × 256 grid (Hohl, 1971). From a local stability analysis done by Toomre (1964), he expected that the short wavelength axisymmetric perturbations could be stabilized by the effects of velocity dispersions. Thus, he added random velocities, with an rms velocity dispersion chosen to be 17% of the circular velocity at the edge of the cold balanced disk; i.e., Q = 1. Instead of the wild behavior found in figure 3-1, a bar developed in less than two rotations. The final configuration was a dense bar-shaped

Figure 3-1: A cold Toomre disk calculated on a 128 × 128 grid.
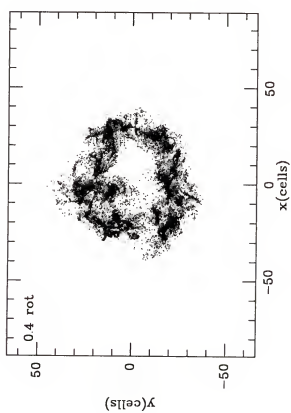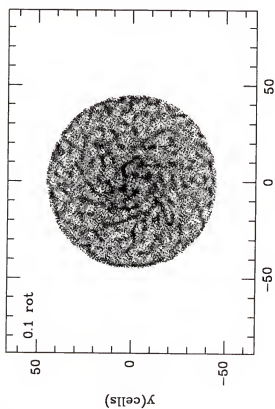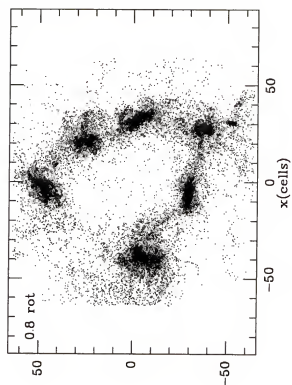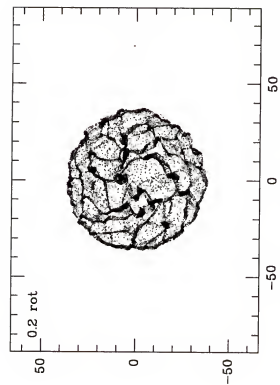
core surrounded by a sparse, nearly axisymmetric disk with large random velocities. This bar instability had not been predicted by the local analysis. Hohl confirmed that the strong instability to bar formation was also present in differentially rotating disks. He was able to build stable, axisymmetric differentially rotating disk but found them to have extraordinarily large random motions, with Q values up to 4. His work suggested that the difference between these stable disks and the bar models lay in the amplitude of the velocity dispersions; smaller dispersions stabilized a cold disk to small scale violent axisymmetric perturbations while disks with dispersions substantially in excess of this were not subject to bar instabilities.

In 1973, Ostriker & Peebles, using a direct N body integrating scheme and working on a truncated Mestel disk, confirmed these findings. In addition, they discovered experimentally the important result that the disk is stable to bar formation if the ratio, t, of the rotational kinetic energy to the absolute value of the gravitational energy is less than 0.14 ± 0.02. This result, although still not yet satisfactorily explained, has proven to be surprisingly useful. "The physics of the bar instability is only indirectly related to this ratio, yet the Ostriker-Peebles criterion provides a surprisingly useful empirical guide for identifying systems that are likely to be stable." (Binney and Tremaine, 1987, p. 374). It has stood the test of time well; to date, there have only been a few cases of stable systems with this ratio greater than 0.14 (Toomre 1981 and references therein).

When attempting to model barred spiral galaxies, an instability is obviously essential. As figure 3-1 shows however, a cold disk is not a useful starting place unless the instabilities can be controlled. It is necessary to have a method of building models in

which the initial instabilities can be managed and constrained. In order for a bar to form, the shorter wavelength perturbations must be damped out. To further stabilize the disk, the longer nonaxisymmetric bar modes must be damped. The Ostriker-Peebles criterion provides clues how to do this. There are two ways to stabilize a disk, quelling bar formation. The first is to add heating in the form of peculiar velocities while decreasing the amplitude of the rotational velocity thereby reducing t from the virial value for a cold equilibrium disk of 0.5 to a smaller value. A second method is to include a radially symmetric halo potential. Such a symmetric potential acts against the development of axisymmetric structure. The use of a halo for this purpose was proposed by Ostriker and Peebles. Both techniques are utilized in the present code.

<u>Halo</u>

When a halo is added, the disk is effectively placed in a deeper potential well, stabilizing it against bar formation. A spherically symmetric halo adds a contribution to the circular velocity field. Although equations (3.4) and (3.5) refer to a disk alone, they suggest that the addition of a halo, which contributes to the disk's rotational velocity, will reduce the length scale of the disk instabilities, thus damping the global instabilities. The halo allows small scale perturbations but stabilizes the longer wavelength perturbations.

The density of the applied halo is given by:

$$
\begin{aligned}
\rho(r) &= \rho_o & 0 \leq r \leq r_o \\
&= \rho_o \left(\frac{r_o}{r}\right)^2 & r \geq r_o \\
&= 0 & r \geq r_e,
\end{aligned}
\tag{3.6}
$$

where r_o is the edge of the constant density core and r_e is the outer edge of the halo. There are a number of advantages of this form of halo; the analytical forces can be calculated easily, it gives rise to a flat rotation curve, and it is physically somewhat realistic. In most of the models the radius of the constant density core of the halo was small, typically 10-20% of the total initial radius in order to represent a bulge component.

The mass can be derived from this density distribution to give, interior to r_o:

$$M_o(r_o) = \frac{4\pi\rho_o r_o^3}{3},\tag{3.7}$$

and within radius r (r > r_o):

$$M_2 = M_o\left[\frac{3r}{r_o} - 2\right] \qquad r_o \leq r \leq r_e$$

$$M_{Total} = M_o\left[\frac{3r_e}{r_o} - 2\right] \qquad r \geq r_e \tag{3.8}$$

The forces associated with the halo are:

$$F_1(r) = -\frac{4\pi G\rho_o r}{3} \qquad r \leq r_o$$

$$F_2(r) = 4\pi G\rho_o r_o^2\left(\frac{2r_o}{3r^2} - \frac{1}{r}\right) \qquad r_o \leq r \leq r_e$$

$$F_3(r) = -\frac{4\pi G\rho_o r_o^3}{3r^2}\left(\frac{3r_e}{r_o} - 2\right) \qquad r \geq r_e \tag{3.9}$$

The contribution to the circular velocity due to the halo is:

$$V_c^2 = \frac{4\pi G\rho_o r^2}{3} \qquad r \leq r_o$$

$$V_c^2 = 4\pi G\rho_o r_o^2\left[1 - \frac{2r_o}{3r}\right] \qquad r \geq r_o$$

$$V_c^2 = \frac{4\pi G\rho_o r_o^3}{3r}\left[\frac{3r_e}{r_o} - 2\right] \qquad r \geq r_e \tag{3.10}$$

Thus the total circular velocity becomes

$$V^2(r) = V_D^2 + V_H^2 \tag{3.11}$$

where D implies disk and H, halo. Plots of the disk and halo contribution to the rotation curve for an n = 0 Toomre disk are shown in figure 3-2 for various values of r_o and r_e. These parameters are given in units of the initial disk radius, 36 cells. The halo contains twice the disk mass within the disk radius. Due to the drop in the halo velocity beyond radius r_e, the edge of the halo was always placed off the edge of the grid. The halo is included to provide an imposed, stabilizing force; the destabilizing effect of this velocity cliff was not explored.
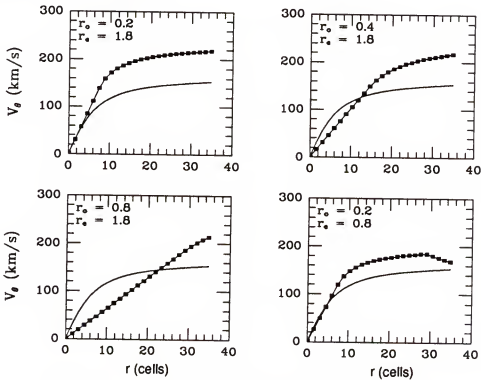
Figure 3-2: Contributions to the rotation curve from the disk (solid lines) and halo (boxes) for various values of r_o and r_e.

The addition of the halo is quite effective at quenching bar formation. Figures 3-3 through 3-5 show a series of cold stellar disks with imposed haloes. The halo mass increases from 1 to 2 to 3 to 5 times the disk mass at the edge of the initial radius. All models were run on a 128 × 128 grid and initialized with a disk mass of 1×10^{11} $M_\odot$, a total radius, R, equal to 36 cells, 12 Kpc, and a softening length, ϵ, of 0.5 cells. Figure 3-6 shows the peculiar radial velocity squared versus disk rotations, at three radii, for the four cases. The stabilizing effect of the halo is evident in all figures. In figure 3-3 the surface density at various rotations are plotted for a model with the halo mass equal to the disk mass at the edge of the initial disk. Although the halo is not massive enough to completely prevent global instabilities, the improvement in behavior over a cold model with no halo is apparent. A bar (approximately 40% of the disk mass) forms in the center of the disk and is surrounded by an extended, featureless, stable disk. In addition to the global bar instability, small-scale instabilities as found in the cold model are seen early in the model. However, the halo does not prevent the dramatic increase in the peculiar velocities at various radii, as seen in the top left plot of figure 3-6. The values plotted are for radii centered at 9 (squares), 19 (circles) and 29 (triangles) cells. This rise in peculiar velocities smears out any spiral features in less than two rotations and prevents any spiral structure that might arise from the motion of the bar through the disk from developing. The bar is apparent in the increased radial velocities in the inner region of the disk in figure 3-6. Another indicator that this model is still fairly unstable is the loss of more than 15 percent of the particles off the edge of the grid in twelve rotation periods.

Figure 3-3: A cold stellar disk with a halo equal to the disk mass.

STARS
$M_{BH}(R) = 1 M_D$
0 rot

0.5 rot

1 rot

1.5 rot

128cool.out87

y(cells)

x(cells)
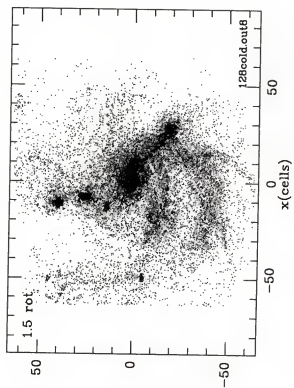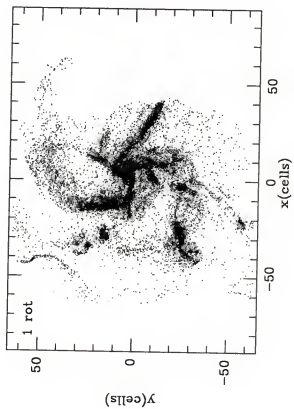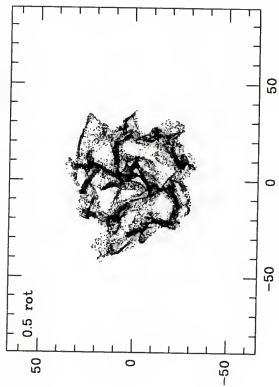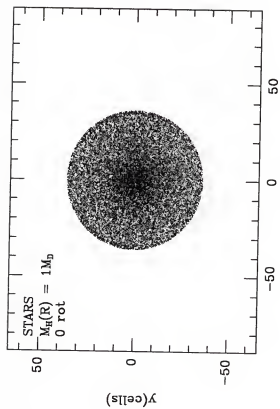
Figure 3-3: continued
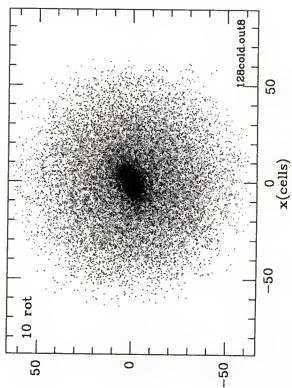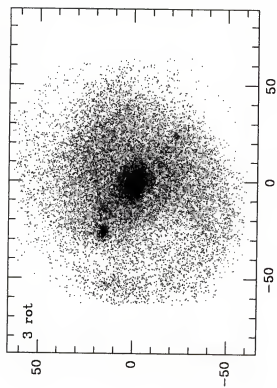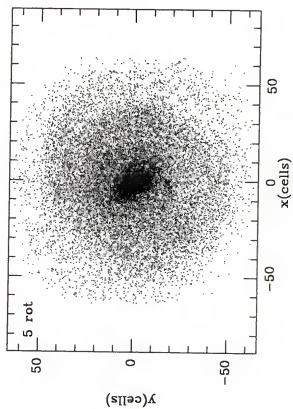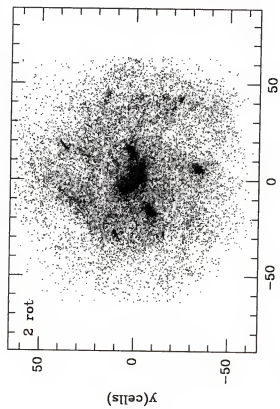
When the halo mass is increased to 2 M_D no bar forms, although the central density is enhanced over the initial configuration. Figure 3-4 shows the surface densities at 0.5, 1, 2 and 10 rotations. The initial disk is that shown in figure 3-3. The denser central region contains approximately 30% of the disk mass. The disk is stabler than the lower halo mass model at all stages of development. In addition, the noncircular velocities, seen in the top right corner of figure 3-6, are significantly lower, especially for the inner disk (r = 9 cells, symbolized by squares). Despite the decrease in noncircular velocities as the halo mass increases, there is still no evidence of stellar spiral structure in the heavier halo mass models. This is because as the halo mass increases, the longer wavelengths perturbations are stabilized and only shorter and shorter instabilities are permitted. This leads to smaller scale spiral features which need only small noncircular velocities to be destroyed.

Figure 3-5 shows the surface density for the final two cases, with halos of three and five times the disk mass. The lower mass model is shown at 1 and 10 rotations in the top two panels of the figure with the higher mass case, at the same rotations, directly below. Even at one rotation, the disks are fairly stable and there is no evidence of either a bar or even a central density enhancement at 10 rotations. The bottom two plots in figure 3-6 are the peculiar radial velocities squared for these cases. It is apparent that as the halo mass increases, the noncircular velocities in the disk decrease. Further evidence of stability is that for $M_H = 3$ M_D, less than 0.2% of the stars have moved beyond the grid boundary, while, for the higher mass case, no particles are lost in 10 rotations.

Figure 3-4: A cold stellar disk with a halo of twice the disk mass.

STARS
$M_H(R) = 2M_D$
0.5 rot

1 rot

3 rot

10 rot

x(cells)

y(cells)
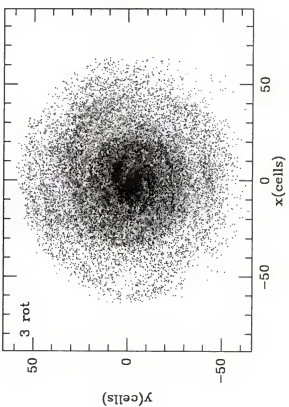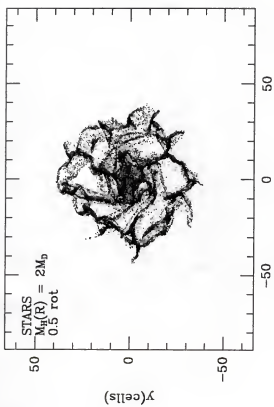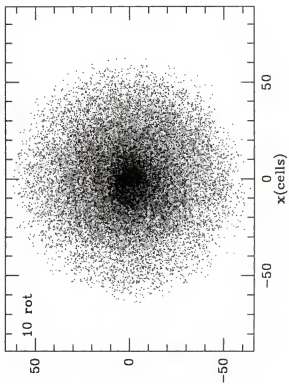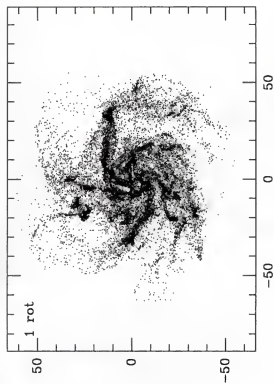
Figure 3-5: Cold stellar disks with haloes of three (top) and five (bottom) times the disk mass.
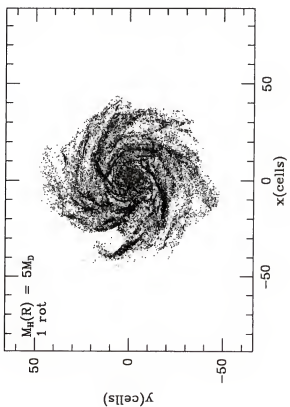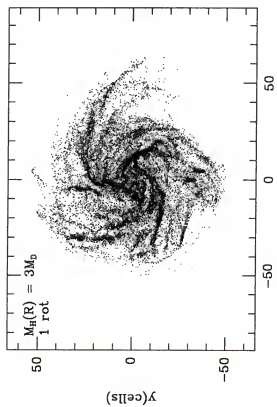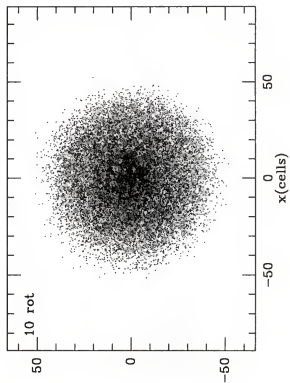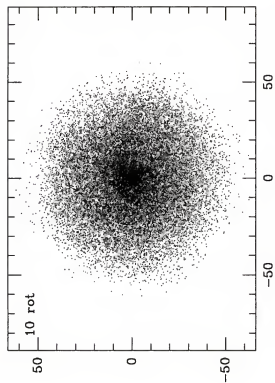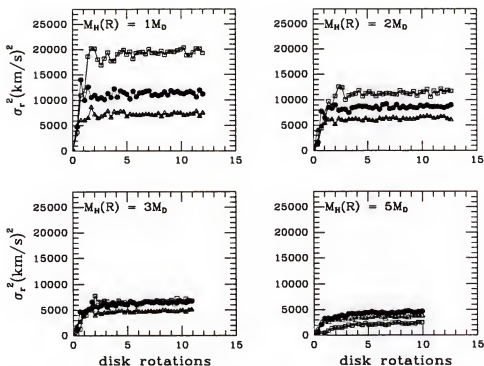
Figure 3-6: Peculiar radial velocities squared for four values of the halo mass

In figure 3-7 the dispersion velocities are compared to the circular velocity for the model seen in figure 3-5, with a halo mass 2 times the disk mass. The ratio of the two is approximately one-third to one-fourth, in good agreement with the observed ratio in our galaxy.

It is evident that a halo alone is capable of completely suppressing bar formation. The masses required to stabilize a model are not unreasonable. In 1990, Persic and Salucci used rotation curves to calculate the disk to halo mass ratio for 58 spiral galaxies (at the optical radius, 3.2 disc length-scales). They found halo masses ranging from 0.19 M_D to greater than 11 M_D. However, the ratio was greater than half for 42 galaxies (72%) of the sample. Therefore, the values needed to stabilize the model disks and
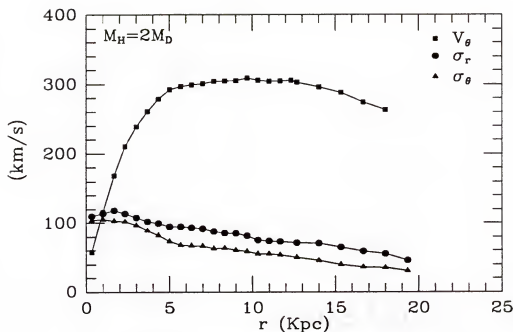
Figure 3-7: Noncircular and circular velocities for a
model with a halo twice as massive as the disk included.

reduce the peculiar velocities to values in accord with observations, one to three times
M_D, are not excessive, but values of one or less would be more representative of the
majority of disk galaxies.

A point that must be emphasized is that the halo forces are calculated analytically
and imposed each time step. Hence they are not incorporated in a self-consistent manner.
This is the rule in most N-body simulations, and little work has been done to establish
whether this lack of self-consistency is important. The lone exceptions are the models
of Hohl (1978) and Sellwood (1980), who modeled the spheroidal mass component as a
population of stars. Sellwood concluded "that the simplifying assumption of a fixed halo
field is quite good and that interactions between the stellar populations do not affect the

stability of the disk." (Sellwood, 1981, p. 363). This result has not been tested in the current code and it does not seem obvious. As Toomre (1977, p. 469) remarks: "It is doubtful, however, whether any postulated haloes, especially if they rotate to some extent, can be relegated safely to an altogether passive role... some two-stream instabilities might well arise between such systems and the cooler disks embedded within them."

The effect of varying the value of r_o will be discussed at the end of the chapter.

Heating

An alternative method of stabilizing a cold disk is to add velocity dispersions while decreasing the circular rotational velocity. This decreases the value of t into the stable regime. In 1964, Toomre determined the minimum radial velocity dispersion necessary to prevent local axisymmetric ($m = 0$) instabilities in a flattened disk. He used a local epicyclic approximation, assuming that the velocity dispersions were small. His result, the Toomre dispersion criterion, is given as:

$$\sigma_{r,min} = \frac{3.36 G \mu}{\kappa},$$ (3.12)

where

μ = the surface density of the disk.

κ = the epicyclic frequency.

The ratio of the radial to the azimuthal dispersion can then be calculated following Chandrasekar (1942) or from the zeroth order Boltzmann equation. This is given by:

$$\frac{\sigma_\phi^2}{\sigma_r^2} = \frac{1}{2}\left[1 + \frac{r}{V}\frac{dV}{dr}\right]$$ (3.13)

where V is the cold rotational velocity (Toomre, 1964).

This method of calculating heating, named after Toomre, has been widely used since its introduction. However, the amount of heating suggested by equation (3.12) will not prevent global bar instabilities. Numerical experiments indicate that this equation must be increased, i.e., multiplied by a "Q" factor, in order to quell global instabilities and construct a completely stable disk. Global stability was found to be guaranteed only if Q was between 2 to 4. The velocity dispersions calculated using such values of Q are not small corrections to the rotation curve; in the center of disks they are many times greater than the rotational velocities. Thus the Toomre formalism is not strictly consistent. In the present code, the Toomre criterion is not used. Instead a new method of calculating velocity dispersions that guarantees global stability was developed.

The method of calculating velocity dispersions is developed by taking moments of the collisionless Boltzmann equation. For an axisymmetric disk, the results may be combined to yield an ordinary differential equation for the radial velocity dispersion as a function of the radius. In many cases this equation must be solved numerically. However, there are a few simple but useful cases that may be solved analytically.

Using standard notation in cylindrical coordinates, assuming axial symmetry $\left(\frac{\partial}{\partial \phi} = 0\right)$, two dimensions (z = 0), the collisionless Boltzmann equation can be written:

$$\frac{\partial f}{\partial t} + \frac{\partial r}{\partial t}\frac{\partial f}{\partial r} + \frac{\partial V_r}{\partial t}\frac{\partial f}{\partial V_r} + \frac{\partial V_\phi}{\partial t}\frac{\partial f}{\partial V_\phi} = 0. \tag{3.14}$$

Letting Q be a velocity dependent function, (note, this is not the Toomre dispersion Q)

a moment of the collisionless Boltzmann equation is:

$$\frac{\partial}{\partial t}\int Qf d\tau_v + \frac{\partial}{\partial r}\int QV_r f d\tau_v + \int f\left(\frac{V_\phi^2}{r} - \frac{\partial\Phi}{\partial r}\right)\frac{\partial Q}{\partial V_r}d\tau_v - \\ \frac{1}{r}\int fV_r\frac{\partial(QV_\phi)d\tau_v}{\partial V_\phi} = 0,$$

(3.15)

where Φ is the gravitational potential, and $d\tau_v = dV_r dV_\phi$. The distribution function, f, is defined to be such that $\int mf d\tau_v = \nu(r,t) \equiv \nu$, the surface density, where m is the mass of each particle. To generate the first equation, Q is set equal to mV_r yielding the conservation of radial momentum:

$$\frac{\partial}{\partial t}\left(\nu\bar{V}_r\right) + \frac{\partial}{\partial r}\left(\nu\bar{V}_r^2\right) + \frac{\partial}{\partial r}(\nu\sigma_r^2) - \nu\left(\frac{\bar{V}_\phi^2}{r} - \frac{\partial\Phi}{\partial r}\right) + \frac{\nu}{r}(\sigma_r^2 - \sigma_\phi^2) \\ + \frac{\nu}{r}\bar{V}_r^2 = 0,$$

(3.16)

where $\bar{V}_r$ is the average radial velocity and $\bar{V}_\phi$ the corresponding azimuthal velocity. It is assumed that the distribution of peculiar radial velocities u_r is symmetric with respect to the local standard of rest, and may be characterized by the radial velocity dispersion, σ_r. At each value of u_r the initial distribution of peculiar tangential velocities is symmetric about the mean, $\bar{V}_\phi$; the corresponding initial tangential velocity dispersion is σ_ϕ. Thus, if the disk is broken up into a large number of concentric rings, one of the axes of the initial velocity ellipsoid is oriented parallel to the tangential direction, and the initial stress tensor is diagonal, containing only the radial and tangential "pressure" terms, $p_{rr} = \nu\sigma_r^2$ and $p_{\phi\phi} = \nu\sigma_\phi^2$, for each annulus. Thus, V_r^2 has been replaced with $\bar{V}_r^2 + \sigma_r^2$ and V_ϕ^2 by $\bar{V}_\phi^2 + \sigma_\phi^2$.

A cross moment equation is formed by letting $Q = mV_rV_\phi$. This yields:

$$\frac{\partial}{\partial t}\left(\nu\bar{V}_r\bar{V}_\phi\right) + \frac{\partial}{\partial r}\left[\nu\bar{V}_\phi\left(\bar{V}_r^2 + \sigma_r^2\right)\right] - \frac{\nu\bar{V}_\phi}{r}\left(\bar{V}_\phi^2 + 3\sigma_\phi^2\right) +$$
$$\nu\bar{V}_\phi\frac{\partial\Phi}{\partial r} + \frac{2\nu\bar{V}_\phi}{r}(\bar{V}_r^2 + \sigma_r^2) = 0. \tag{3.17}$$

Multiplying equations (3.16) by $\bar{V}_\phi$ and subtracting this from (3.17) gives:

$$\nu\bar{V}_r\frac{\partial}{\partial t}\left(\bar{V}_\phi\right) + \nu\bar{V}_r^2\frac{\partial\bar{V}_\phi}{\partial r} + \frac{\nu\bar{V}_\phi\bar{V}_r^2}{r} - \frac{2\nu\bar{V}_\phi}{r}\sigma_\phi^2 + \nu\sigma_r^2\left(\frac{\partial\bar{V}_\phi}{\partial r} + \frac{\bar{V}_\phi}{r}\right) = 0. \tag{3.18}$$

Reviews of this approach may be found in Mihalis and Routley (1968), Binney and Tremaine (1987) and references therein.

For a time independent model with strictly circular average velocities $\bar{V}_r = 0$, and equation (3.16) becomes:

$$\frac{d}{dr}(\nu\sigma_r^2) + \frac{\nu}{r}(\sigma_r^2 - \sigma_\varphi^2) = -\nu\frac{\partial\Phi}{\partial r} + \frac{\nu}{r}\bar{V}_\phi^2 = -\frac{\nu}{r}(V_o^2 - \bar{V}_\phi^2) \tag{3.19}$$

where $V_o(r)$ is the rotational velocity of a cold disk. $\bar{V}_\phi(r)$ is assumed to have the same form as $V_o(r)$ but a smaller amplitude; $\bar{V}_\phi(r) = kV_o$. The dimensionless constant **k** ($0 \leq$ **k** ≤ 1) parameterizes the heating. For **k** = 1, the disk is cold with no velocity dispersions, while **k** = 0 corresponds to a disk supported entirely by internal pressure. Substituting these expressions into equation (3.19) yields the result:

$$\frac{d}{dr}(\nu\sigma_r^2) + \frac{\nu}{r}(\sigma_r^2 - \sigma_\varphi^2) = -\nu\frac{\partial\Phi}{\partial r} + \frac{\nu}{r}k^2V_o^2 = -\frac{\nu V_o^2}{r}(1 - k^2). \tag{3.20}$$

With the same restrictions, equation (3.18) reduces to:

$$\sigma_\phi^2 = \frac{\sigma_r^2}{2}\left(1 + \frac{r}{V_o}\frac{dV_o}{dr}\right), \tag{3.21}$$

exactly the result given in equation (3.13). Equation (3.21) is quoted explicitly in this form by Toomre (1964). Although this expression is well known from epicyclic theory, it is important to realize the generality of the results in the present context, where the initial orbits are constrained to be circular. In particular, the velocity dispersions need not be small, as is the case in linear epicyclic theory. In particular the asymmetric drift, the difference between the average velocity and the cold circular velocity, is known and equals $(1 - k)V_o$.

Inserting equation (3.21) into equation (3.20) yields a first order, ordinary differential equation for $\sigma_r{}^2$:

$$\frac{d}{dr}\left(\nu\sigma_r^2\right) + \frac{\nu}{2r}\sigma_r^2\left(1 - \frac{dlnV_o}{dlnr}\right) = -\frac{\nu\left(1 - k^2\right)V_o^2}{r}. \tag{3.22}$$

Having specified k and a form for the surface density and cold rotation field, this equation can be solved for the radial velocity dispersion. As a boundary condition, $\sigma_r{}^2$ approaches zero at the edge of the disk. In many cases, equation (3.22) must be solved numerically. The result is shown graphically in figure 3-8 for a correctly truncated Toomre n = 0 disk with a radius of 12 Kpc and a mass of 1×10^{11} M$_o$. Also shown in the figure is the Toomre dispersion minimum, calculated using Toomre's criterion to suppress local instabilities (Toomre, 1964). The Toomre dispersion minimum lies between curves with k of 0.8 and 0.9.

For a correctly truncated Toomre n = 0 disk, equation (3.22) must be solved numerically. If isotropic heating is imposed, $\sigma_r = \sigma_\phi$, equation (3.20) simplifies to
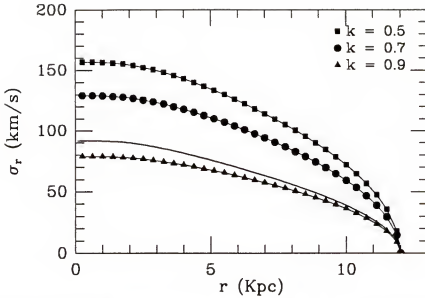
Figure 3-8: The radial dispersion as a function of radius for an exact truncated Toomre disk. Solid line shows the Toomre dispersion minimum.

the form:

$$\frac{d}{dr}(\nu\sigma_r^2) = -\nu r\left(\Omega_o^2 - \Omega^2\right) = -\nu r\Omega_o^2(1 - k^2). \tag{3.23}$$

Figure 3-9 shows the surface density distribution at various rotations, for this isotropic, correctly truncated disk with a **k** of 0.5. The disk has expanded slightly from its initial configuration but is stable, displaying no evidence of either bar or spiral instabilities.

For an infinite Toomre n = 0 disk, if isotropic heating is imposed, equation (3.23) may be integrated analytically to yield:

$$\sigma_r^2(r) = C_1^2(1 - k^2)\frac{\sqrt{b^2 + r^2}}{b}\left[\ln\left(\frac{\frac{b + \sqrt{b^2 + r^2}}{\sqrt{b^2 + r^2}}}{\frac{b + \sqrt{b^2 + R^2}}{\sqrt{b^2 + R^2}}}\right)\right], \tag{3.24}$$

where the infinite density distribution has been truncated at r = R. A plot of this curve for various values of **k** is shown in figure 3-10 for R = 24 Kpc and $M_D = 1 \times 10^{11}$ M_o. As the heating values can be solved for analytically rather than numerically, it is

Figure 3-9: A correctly truncated Toomre disk with $k = 0.5$.
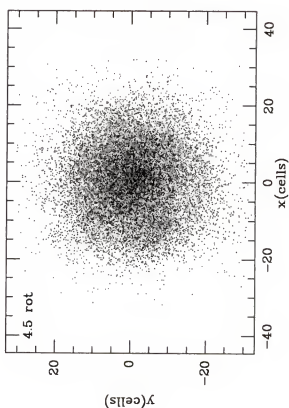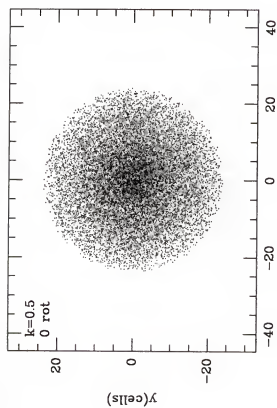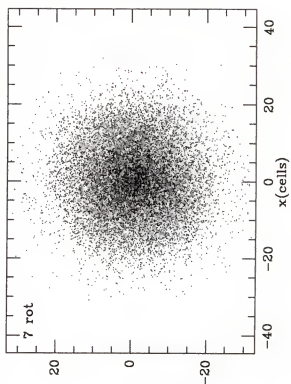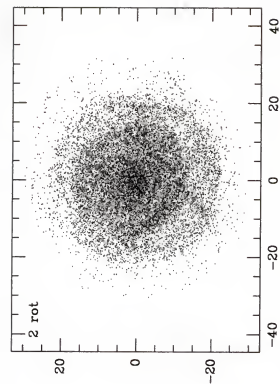
easy to work with and easy to vary the heating across the disk. This is an important feature when trying to form bar and spiral features as will be discussed in Chapter 4.
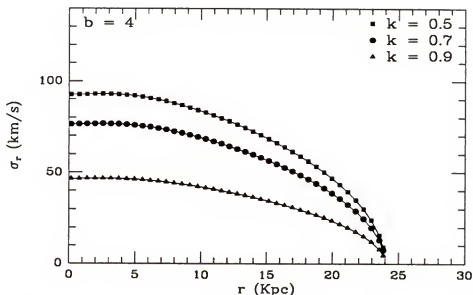
Figure 3-10: The radial velocity dispersion as a function of radius, for an infinite Toomre disk.

In figure 3-11, the stellar surface densities after 4.0 rotations for values of **k** ranging from 0.9 to 0.6 are presented. The stabilizing effects of the peculiar velocities can clearly be seen as **k** decreases. The figures get less bar-like, lose fewer particles off the edge of the grid, and show less wandering of their center of mass as **k** decreases.

The Ostriker-Peebles criterion, that a **t** value of ≤ 0.14 is necessary to stabilize a disk implies stability should be guaranteed if **k** is reduced to 0.53. Figure 3-12 shows a sequence of density plots for **k** = 0.5 at 0, 4, 8 and 12 disk rotations. The value of the softening length, ϵ is 0.5 cells. Radial velocity information for this model is displayed in figure 3-13. In the top left plot the dispersions squared versus disk rotations for three

Figure 3-11: Stellar disks after 4 rotations for values of **k** ranging from 0.9 to 0.6.
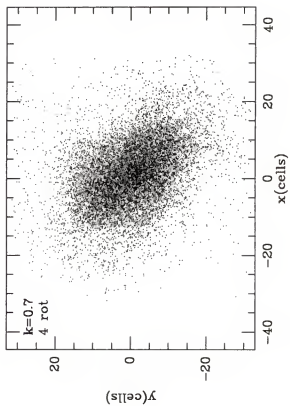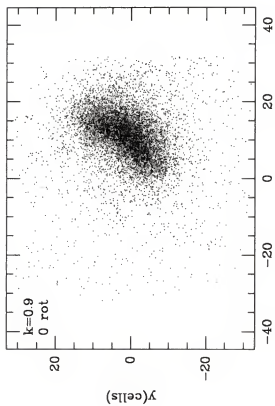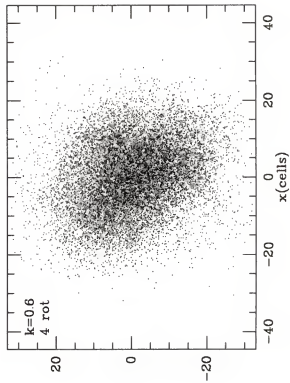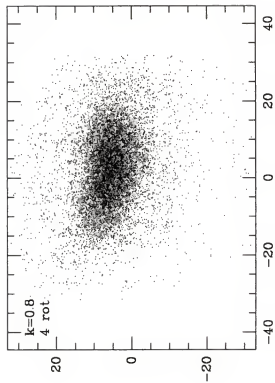
Figure 3-12: A stable stellar disk with a **k** of 0.5.
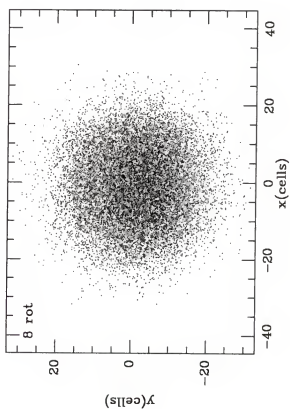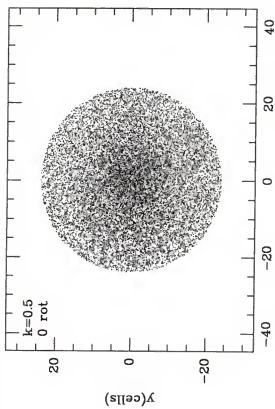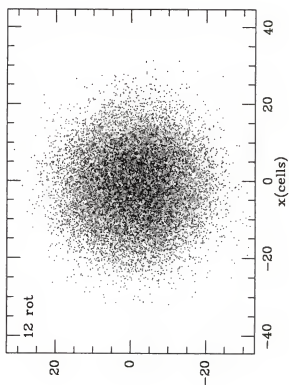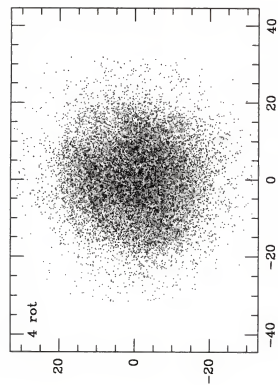
values of radius, 4.5, 9.5 and 14.5 cells, are shown. Values decreases from the inner
radii (boxes) to the outer (triangles). The curves display a steep rise initially (within
the first two rotations) but then settle down. The final value is close to the initial at the
outer edge of the disk but higher in the inner disk. The top right graph in figure 3-13
shows the initial and final rotation curves. In the bottom left, the initial and final radial
peculiar velocity versus radius are displayed and the initial and final mass distributions
are presented at the bottom right. The disk has readjusted slightly but appears to have
reached a stable state not significantly different from the initial disk. The total change
in angular momentum after 14 disk rotations is -0.36%
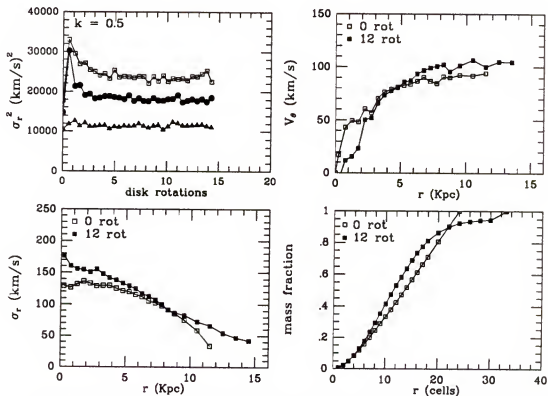
Figure 3-13: Velocity and mass fraction plots for a stable disk with a **k** of 0.5.

Isotropic heating was imposed when initializing this disk; to check if the dispersions maintained isotropy, the radial and tangential peculiar velocities are plotted in figure 3-14 for two values of the radius, 4.5 and 14.5 cells. The curves from the inner radius are on top, the outer, below. In the interior region of the disk, the two curves diverge initially but then rejoin and remain approximately equal after a few rotations. This is expected from equation (3.21), which predicts the equality of the two dispersions for solid body rotation. As figure 3-13 shows, in the inner region of the disk, solid body rotation is a good approximation. In the outer regions however, the tangential component is consistently lower than the radial, also in accord with equation (3.21). The rotation curve is rising at this radius, and the tangential component should be more than half the radial. The ratio of the two is approximately 0.78.
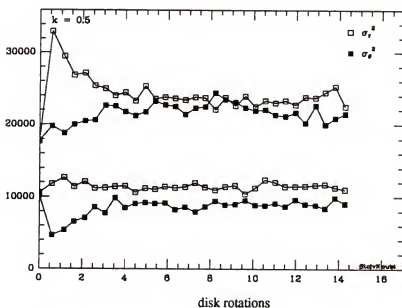
Figure 3-14: Radial and tangential peculiar velocities squared for r = 4.5 (top) and 14.5 (bottom) cells.

To apply the heating to a Toomre disk, the disk is broken up into rings, typically 80 to 100. The radial velocity dispersion is calculated at the center of each annulus. For the isotropic heating, the total dispersion is $\sqrt{2}$ times the radial dispersion. This value is calculated and applied to each particle in the ring at a random angle. The initial distribution of the total noncircular velocity can be seen in the top left of figure 3-15. The model shown is that presented in figures 3-12 and 3-13. The number of
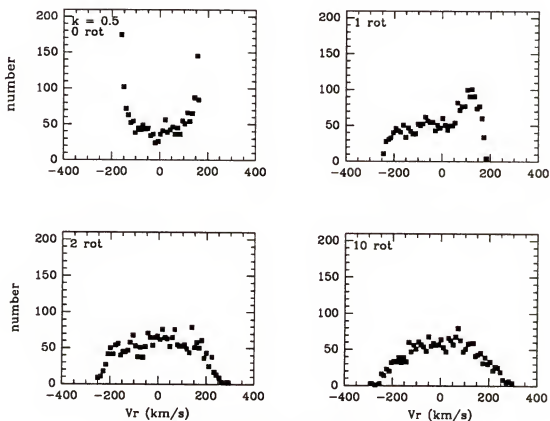
Figure 3-15: Radial velocity distribution, for k = 0.5

particles rises sharply at plus and minus the actual value of the dispersion. However, this distribution is not maintained, after two rotation periods the distribution has become

roughly Maxwellian. It retains this shape through subsequent model development. This initial adjustment is believed to be one cause of the steep jumps in the peculiar velocities found in the top left of figure 3-13.

It is clear from the results above that the method of heating adopted stabilizes a disk if **k** is decreased to approximately 0.5. Although the values of σ_r shown in figure 3-10 for **k** seem high, especially in the outer regions of the disk, it is important to realize that they scale with the radius for a given mass galaxy. In figure 3-10, the radius was 24 Kpc, if it is increased to 30 Kpc, the value of σ_r at the innermost radius falls from 93 to 83 km/s. More relevant numbers are the ratio of the dispersions to the rotational velocities or the Q value. The latter is misleading in this case because the rotational velocities are decreased by a substantial amount as **k** decreases, giving reasonable Q values even when the dispersions are greater than the circular velocity. The most reliable estimate of Q in a real galactic disk comes from the solar neighborhood and is given as 1.7, with a probable range of 1 to 3 (Binney and Tremaine, 1987). There is also new evidence that the noncircular velocities in our galaxy are larger than first suspected, and may reach close to 100 km/s near the galactic center (Lewis and Freeman, 1989). In the solar neighborhood, the rms random velocity of the late type dwarfs is about 60km/s (Binney and Tremaine, 1987). A **k** of 0.5 gives initial dispersions greater than the rotational velocities over much of the disk. This is unrealistic and is made more so by the fact that the dispersions increase over their initial values, especially at the disk interior. It would appear that the velocity dispersions needed to completely stabilize a simulated disk are well in excess of observed dispersions in our galaxy. This is the case for most numerical

studies. There are a number of ways to decrease the heating; a halo can be included or the stars can be cooled by accretion. Sellwood and Carlberg (1984) experiment with accretion and star formation as methods of cooling a stellar disk. They were able to cool the disk enough to maintain stellar spiral structure for more than 12 rotation periods.

If a halo equal in mass to the disk is included in a model, the disk can be stabilized with a k of approximately 0.7. Figure 3-16 shows a stable Toomre $n = 0$ disk with a halo equal to the disk mass and $k = 0.7$ at 0, 4, 8 and 16 rotations. The velocity dispersions after 16 rotations, shown in figure 3-18, are significantly lower than the 0.5 results. In addition to the fact that the addition of a halo has increased the rotational velocity, the model gives a much more reasonable (one-half to one-third over the outer disk), but still high, ratio of peculiar to circular velocities. For this model r_0 was equal to 0.4 R.

The effectiveness of the halo at stabilizing a disk depends mainly on the ratio of the halo to disk mass within the disk radius, but also varies to a lesser extent with the value of r_0. A series of models was calculated with $k = 0.7$ and a halo equal to the disk mass but with varying values of r_0. The resulting surface densities, at 12 rotations, are seen in figure 3-17, and peculiar velocities squared in 3-18, for three values of r_0. The value of r_0 is 0.2 R in the first model, 0.4 in the second, and 0.8 in the third. When r_0 changes, it is not possible to get identical values for the total halo mass within the grid. In the first three models, the halo mass within R is equal to the disk mass, but the total halo mass within the grid increases from 2.0 to 2.13 to 2.73 M_D. Although the mass of the halo beyond the disk outer radius is not expected to affect the model, the disk does expand somewhat, resulting in a final radius that extends beyond the initial. To ensure

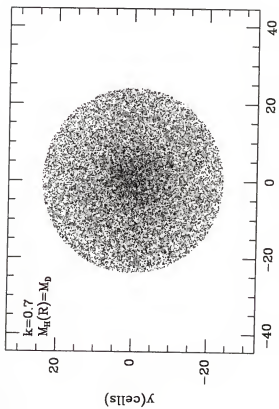Figure 3-16: A stable stellar disk with a k of 0.7 and a halo equal in mass to the disk mass.
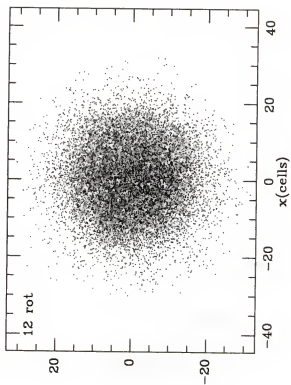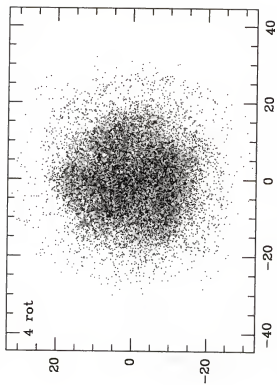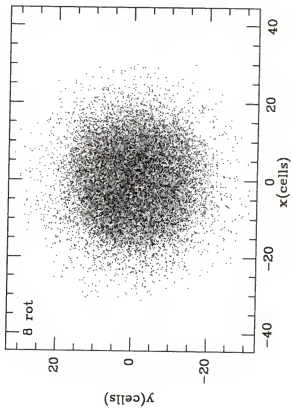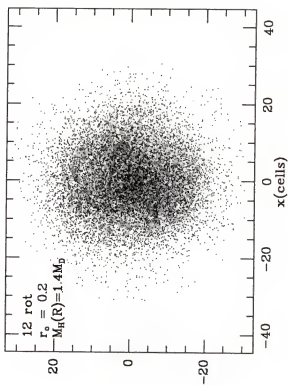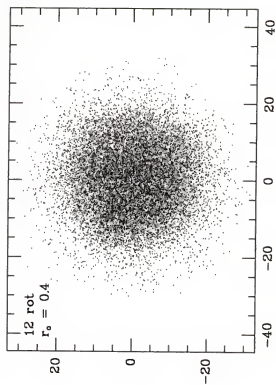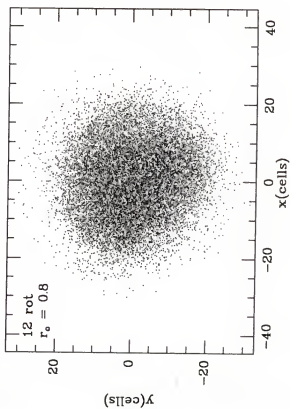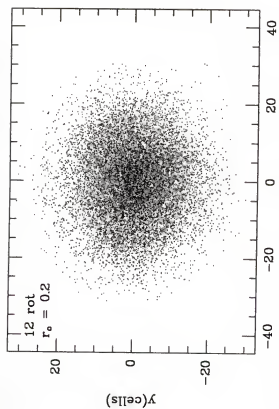
Figure 3-17: The stabilizing effect of increasing r_o for a stellar disk with a k of 0.7 and a halo mass equal to the disk mass.

that the effects seen in figures 3.17 and 3.18 are due to the value of r_o rather than the increased halo mass on the grid, model 4 was calculated, with a total halo mass of 2.73 M_D, but r_o equal to 0.2. This gives a halo mass of 1.42 M_D within R. Density plots of the four cases at twelve rotations show a more enhanced central density for the lowest value of r_o. In addition, the velocity dispersions decrease fairly significantly in the inner regions of the disk as the value of r_o increases. The number of stars lost
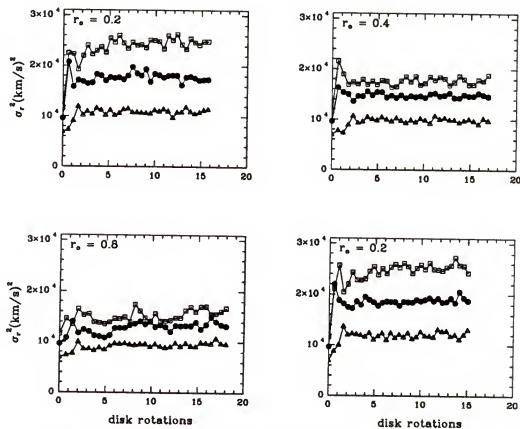
Figure 3-18: Peculiar radial velocities squared for different values of r_o

varies from 8.5 to 3.4 to 1.8 to 6.7%. The differences between the r_o = 0.4 and 0.8 models seem minor, but both give better results than the r_o = 0.2 case. However, in

two-component models in which the heating varies across the disk (discussed in the next chapter), this straightforward behavior is not found. Models with different values of r_o showed substantially different development and haloes with smaller values of r_o were more successful at stabilizing the disks.

Sellwood (1981) did tests with bulges and found faster bar formation occurred as the bulge was made more centrally concentrated. In addition, more central concentration of the bulge leads to a bar which is shorter and rounder. Bardeen (1975) found that the halo was most effective when the halo density was less than the disk density in the interior. He also found that the less centrally condensed the halo, the better it stabilized the disk, in agreement with the above results.

Stabilizing a Gaseous Disk

As with cold stellar disks, cold gaseous disks are violently unstable. Figure 3-19 shows the evolution of a cold Toomre n = 0 gas disk. ($\alpha = 0.1$, $\sigma = 0.6$ cells). The final configuration does not differ significantly from the cold stellar case. In order to construct useful models that include a gas component, the cold gas disk must also be stabilized. However, unlike the stellar case, little effort was made to create stable gas models with a significant amount of heating. Observations show the gas in the interstellar medium to be very cold, with peculiar velocities in the range of 3-8 km/s for large clouds.For this reason, the gas heating was kept small. However, as the gas is only a small percent by mass of the total mass, it is expected to follow the stellar potential and so, stellar heating

acts to stabilize a gas disk. If a cold gas component, 10% by mass, is run with a stable stellar disk with a $k = 0.5$, the stellar disk will act to stabilize the cold gas component.

In addition, a halo can be used to stabilize the cold gas component. Using only a halo as a stabilizing agent, figure 3-20 shows that a halo of mass 3 times the disk easily quells bar formation in a gaseous disk. However, even a cursory glance shows how different the gas model is from the stellar halo models (figures 3-3, 3-4 and 3-5). Due to the dissipative nature of the viscous force, the gas component has maintained spiral structure through 12 rotations, whereas all hints of spiral structure in the stellar component are gone within about 3 rotations.

In this chapter, methods of preventing the instabilities that destroy a cold disk were presented. Small amounts of heating can stabilize the violent, short wavelength, axisymmetric perturbations without eliminating the bar instability. Increasing the peculiar velocities beyond this point, will damp out the longer wavelenthg bar instability, leaving a stable, featureless disk. A halo eliminates the global bar modes, the more massive the halo, the shorter the remaining perturbations. In the next chapter, methods of manipulating these instabilities in a way that leads to the formation of realistic looking bar and spiral features will be presented.

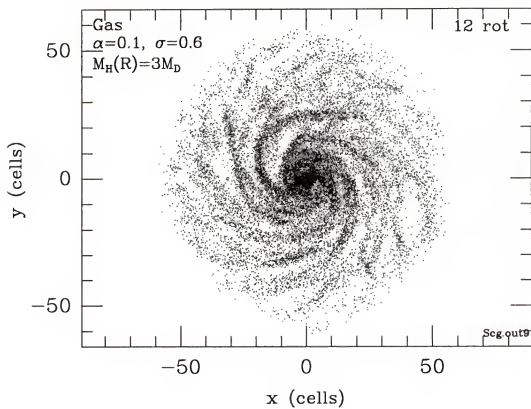Figure 3-19: A cold Toomre gas disk.
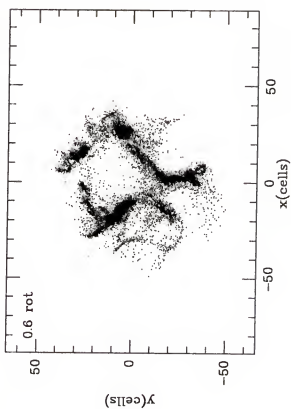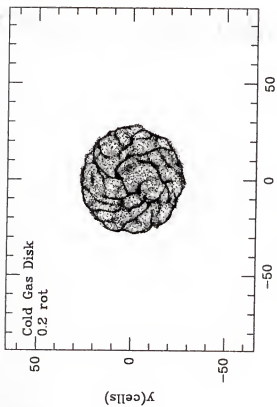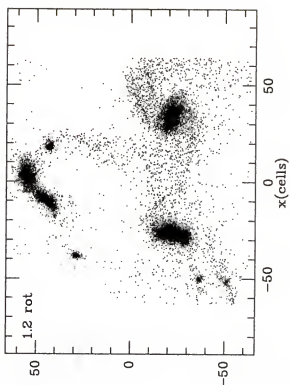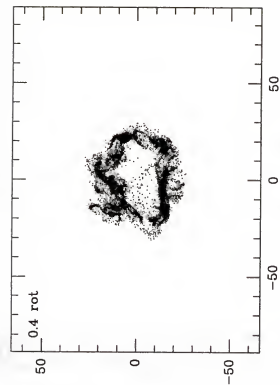
Figure 3-20: A cold gas disk stabilized by a halo 3 times the disk mass.

CHAPTER 4
NUMERICAL MODELING

The procedure followed to produce two component models of barred spiral galaxies is outlined below. Before models with structure can be formed, a method to stabilize cold disks must be developed. As discussed in the preceding chapter, this is be achieved by adding a halo, and/or heating in the form of peculiar velocities while decreasing the circular velocity. Bar and spiral formation can then be induced in a variety of ways. All models presented were calculated on a 128×128 grid.

Cold Disks With Haloes

The first method of inducing structure is to start with a cold, two-component disk with a halo. A halo of modest mass will prevent the disk from becoming violently unstable, while a more massive one will completely inhibit bar formation. In the previous chapter, it was shown that for cold stellar models with haloes, the stars heat up as the model progresses. As their peculiar velocities increase, the stellar spiral structure disappears, usually within two to three rotations, leaving a stable stellar disk. In similar gas models, however, the behavior is quite different. Due to viscous dissipation, the gas remains cool and local spiral density enhancements are not smoothed out by peculiar velocities. The lower velocity dispersion of the gas allows shorter wavelength pertubations to develop there, yielding structure within individual spiral arms.

Carlberg and Freeman (1985), CF, generate spiral structure in two-component disk models by including relatively massive halos, ranging from 2.5 to 10 times the disk mass at the initial disk radius. They use equal mass stellar and gaseous particles with a ratio of stars to gas of three to one. The gas viscous force follows the method of Schwarz (1981), a simple sticky particle scheme. They reach four major conclusions, which can be summarized as follows:

1. A great deal of spiral structure is found but no bar formation takes place. The gas arms are much narrower than the stellar arms and show more structure.

2. The number of arms is a function of the disk to halo mass ratio. This is in agreement with Toomre's local analysis (Toomre, 1964; Toomre, 1981; Sellwood and Carlberg, 1984) and is also found by Elmegreen (1991).

3. The gas component loses angular momentum to the stellar component.

4. The stars heat up and lose spiral structure. The final value of Q varies from 2.5 to 4.0.

A number of models of this type were calculated and compared with the CF results. The algorithms, particularly the gas components, and models are not identical so this was a good test of both the code, especially the gas component, which has not been widely used, and the CF results, which have not been duplicated. Only general qualitative similarities were sought. Two gas distributions were tested. In the first, the initial gas and stellar distributions were identical Toomre n = 0 disks. More emphasis was put on this case because it is closer to the CF scenario. In the second, the stellar particles are initialized as a Toomre n = 0 disk, but gas particles are placed uniformly throughout the disk. The latter case is closer to physical reality, at least in our galaxy. As Bertin
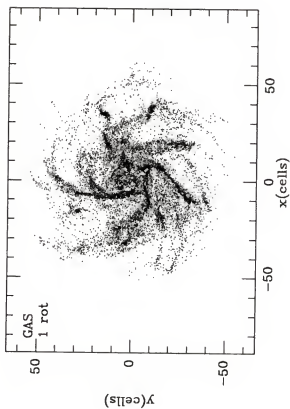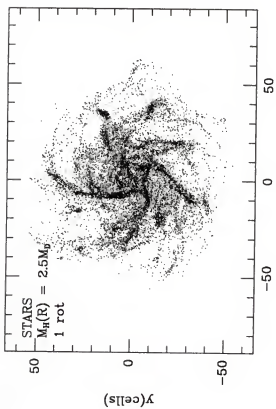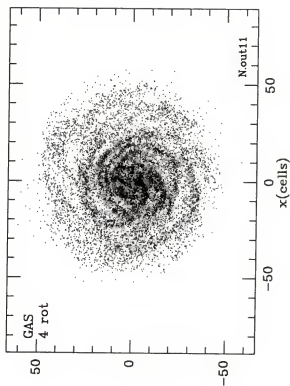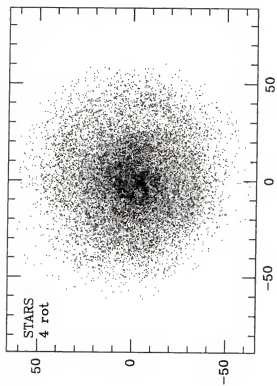
(1990) points out, while the total amount of gas present in spiral galaxies is believed to be less than ten percent even for gas rich spirals, if the local surface density of gas is compared relative to the local stellar density, the gas becomes quite important in the outer part of the disk. In the radial range 3-4h* (h* is the exponential disk scale length), the local value of the ratio of gas to stellar surface density can typically be 50%. This occurs because the stellar disk density drops off exponentially, while the HI distribution is flatter and more extended. Thus, the ratio often increases exponentially with radius. Burton and Gordon (1978) find that the HI in the Milky Way has an approximately flat profile between radii of 5 to 13 Kpc. By letting the stellar distribution fall off while the gas remains constant, an attempt was made in the code to mimic this effect. The number of gas particles per cell initially varies from six to seven depending on the total number of particles and the initial radius. That is enough to ensure that each gas particle has between five and ten near neighbors, if not initially, then as the model evolves. In both distributions the gas component is typically made five to twenty percent of the total disk mass.

The results agree extremely well with the findings of CF. All four of their general conclusions are verified. Figure 4-1 shows density plots of a model with 10% gas by mass and a halo mass of 2.5 $M_D(R)$. Model parameters include; $\alpha = 0.09$, $\sigma = 0.6$ cells, N = 40000, R = 36 cells (12 Kpc). The stellar component is shown in the top panels, with the corresponding gas distributions shown below. The number of disk rotations is labeled in the top left corner of each plot. Both components are initialized as Toomre n = 0 disks. Three of the CF results are verified. No bar forms in either component in this model, or
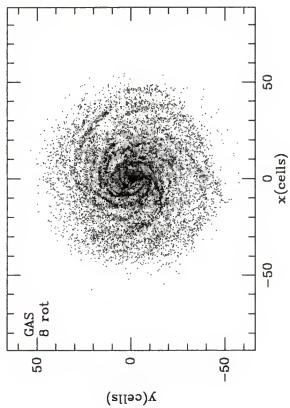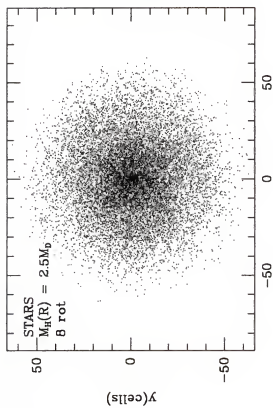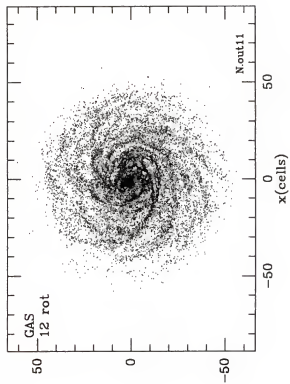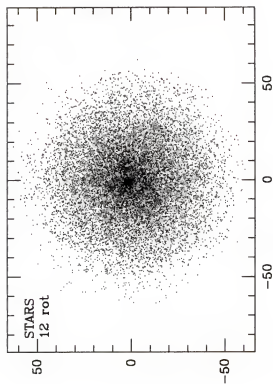
for any model with a halo mass greater than approximately 1.5 to 2 $M_D(R)$. Early in the model, spiral structure is found in both components. However, by three rotations most of the spiral structure is lost from the stellar disk, certainly all the small-scale structure seen in the gas arms is absent. By six rotations, only hints of spiral structure are found in the stellar disk. In contrast, after twelve disk rotations, the gas still shows a great deal of spiral activity. Neither component appears to evolve significantly after four to five rotations. Figure 4-2 shows the stellar and gas peculiar radial velocity squared versus disk rotations for four values of the radius (r = 9 cells, symbolized by boxes, 19 cells; circles, 29 cells; triangles and 42 cells; pentagons). For both components, the initial noncircular velocities are zero. In both, the peculiar velocities rise rapidly within the first one to two rotations. However, the maximum amplitude of the gas dispersional velocity is lower than the stellar and the evolution after two rotations is quite different. In the stellar case, the peculiar velocities are higher in the center and fall off across the disk. They remain fairly constant after two rotations. Due to the massive halo, the stellar dispersions are small with respect to the rotation curve (approximately one third at r = 5Kpc, and one fifth at 10Kpc). The gas noncircular motions however, peak sharply between one and two rotations and then fall off dramatically at all radii due to the viscous forces. In the outer three radii, the gas radial dispersion lies between 20 to 30 km/s. Figure 4-3 shows the peculiar radial velocity versus radius, for both the stars and gas, at 12 rotations, on the left. The gas dispersion is flatter and less than half the magnitude of the stellar component. The rotation curves at the same time are shown on the right. The tangential peculiar velocities exhibit the same behavior as the radial, but

Figure 4-1: A cold two-component model with a halo 2.5 times the disk mass included. Gas makes up 10% of the disk mass. Stars are shown on top, the gas below.

STARS
$M_H(R) = 2.5 M_\odot$
1 rot

STARS
4 rot

GAS
1 rot

GAS
4 rot

N.out11

x(cells)

y(cells)

Figure 4-1: continued

STARS
$M_H(R) = 2.5M_p$
8 rot

STARS
12 rot

GAS
8 rot

GAS
12 rot

N.out11

x(cells)

x(cells)

y(cells)

y(cells)
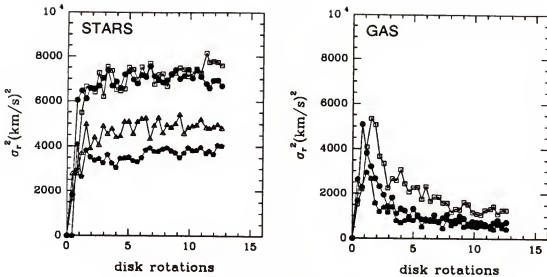
Figure 4-2: Peculiar radial velocity squared for the stellar
(left) and gas (right) component for four values of the radius.

are of lower magnitude, approximately 0.8 the radial dispersion in the inner regions

and 0.4 in the outer.

The third finding of CF, that the gas component loses angular momentum to the

stellar is also verified. After five rotations, the gas angular momentum has decreased

by 5.16% while the stellar has increased by 0.54% (the gas mass is one tenth the total

mass). After 10 rotations, these numbers are −6.58% and +0.67%.

As the halo mass increases with respect to the disk mass, the longer wavelength

perturbations are damped out, allowing for finer spiral structure. This is demonstrated in

the models as an increase in the number of spiral arms with increasing halo to disk mass

ratios. The next sequence of figures illustrates this phenomenon. All models contain

20% gas by mass. Model parameters are; $\alpha = 0.15$, $\sigma = 0.6$, N = 40000 and R = 36

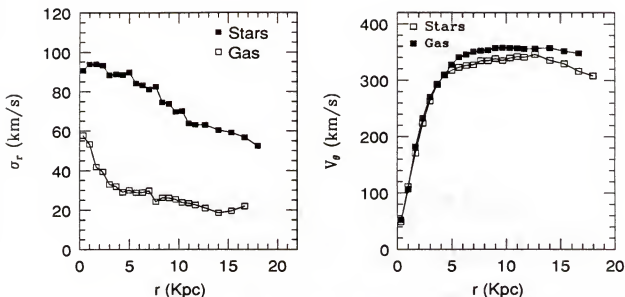cells (12 kpc). The halo mass is increased from 1 (figure 4-4) to 2.5 (figure 4-5) to 5

Figure 4-3: Peculiar radial velocity versus radius and rotation
curves for the stellar and the gas components after 12 rotations.

(figure 4-6) times the disk mass. Again stellar figures are on the top, gas plots below.
All three models are shown at 2, 5, 7 and 10 disk rotations.

The lowest halo mass model, $M_H(R) = 1\ M_D$ (figure 4-4) is the only one in which a
bar develops. The halo is not massive enough to contain the instabilities associated with
a cold disk and the disk expands initially. As the model proceeds, the disk stabilizes
as the peculiar velocities increase, but many particles are lost off the grid. A single
spiral (or ring) feature appears by five rotations and is maintained through a number of
rotations. By 10 rotations however, this feature appears to have spread slightly, into a
more flocculent structure and the number of gas spiral arms has increased. From figures
not shown, the gas does not appear to settle into its final configuration until approximately
8 rotation periods. CF also found an increase in the number of gas arms, with time in
some of their models. It is believed this increase is due to the heating of the stellar disk.

Figure 4-4: The first in a series of cold two-component models with haloes. Gas makes up 20% of the disk. $\alpha = 0.15$, $\sigma = 0.6$ cells. The halo is equal in mass to the disk mass. Notice the increase in the number of gas spiral arms as the model progresses.
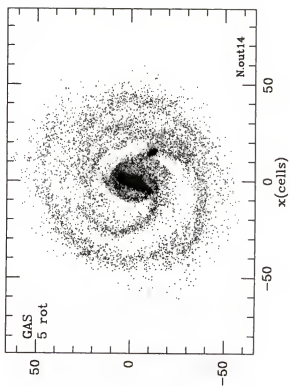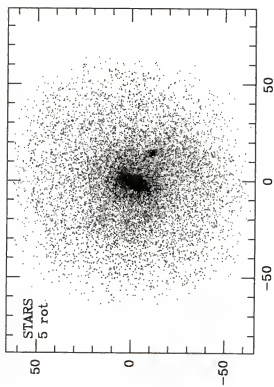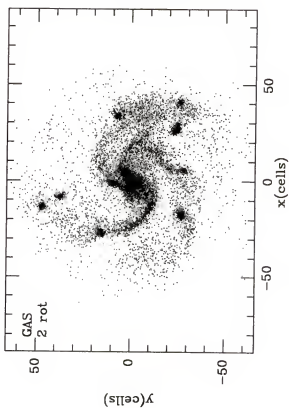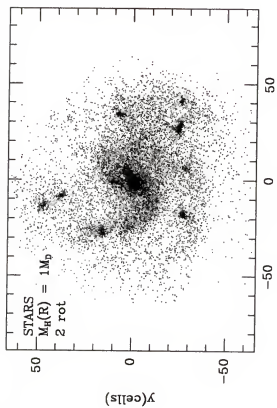
STARS
$M_H(R) = 1 M_D$
2 rot

STARS
5 rot

GAS
2 rot

GAS
5 rot

Nout14

y(cells)

x(cells)

Figure 4-4: continued

STARS
10 rot

STARS
$M_H(R) = 1M_D$
7 rot

GAS
10 rot

N.out14

GAS
7 rot

x(cells)

x(cells)

y(cells)

y(cells)
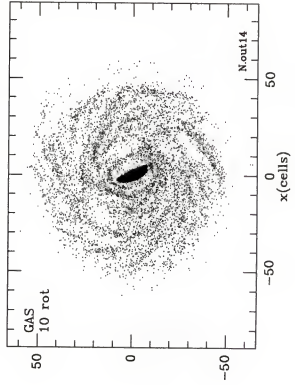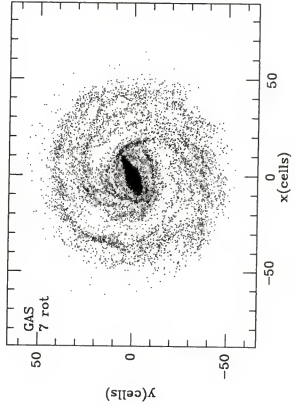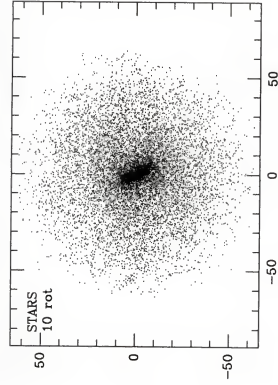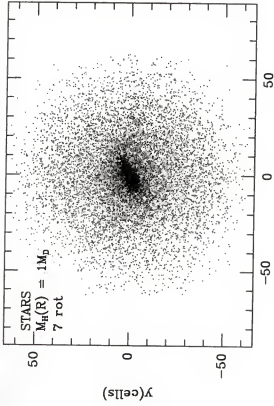
As the disk heats up, it begins to act as a halo component in terms of spiral activity (Carlberg and Freeman, 1985). This increase is seen in all three models but is most prevalent in the lowest halo mass case, the case in which the peculiar velocities rise the most. The final gas configuration shows a small bar surrounded by a disk that displays a great deal of spiral activity. In the stellar component, a small bar is surrounded by a featureless disk.

For $M_H(R) = 2.5 \, M_D$, (figure 4-5) a slight central enhancement, but no bar, forms. Comparing this model to the previous one, the initial behavior is more restrained. By five rotations, the increase in the number of spiral arms over the previous case is obvious. Because fewer particles have been lost off the grid or locked in a central bar, more gas particles inhabit the arms, sharpening and defining their appearance. More importantly, the noncircular velocities in both the stellar and gas component are significantly lower than in the previous model, allowing for finer spiral structure. In the final model, with the most massive halo, (figure 4-6), these trends continue; initially, no violently unstable behavior is displayed and the dispersional velocities are small enough to permit short wavelength perturbations, allowing for intricate, narrow, small scale spiral structure. In addition, little material is trapped in the center or lost off the grid, allowing more particles to participate in the spiral activity. In none of the models does spiral structure persist in the stellar component. Although the stellar heating decreases as the halo mass increases, in all models it is significantly higher than the gas dispersions and in all cases stellar spiral structure is lost by 6 rotations. To summarize, as the halo mass increases, longer wavelength perturbations are damped out, leaving only the smaller-scale perturbations

Figure 4-5: The second in a series of cold two-component models with haloes. Gas makes up 20% of the disk. $\alpha = 0.15$, $\sigma = 0.6$ cells. The halo is 2.5 times as massive as the disk.

STARS
$M_M(R) = 2.5M_0$
2 rot

STARS
5 rot

GAS
2 rot

GAS
5 rot

Nout9

x(cells)

x(cells)

y(cells)

y(cells)

Figure 4-5: continued
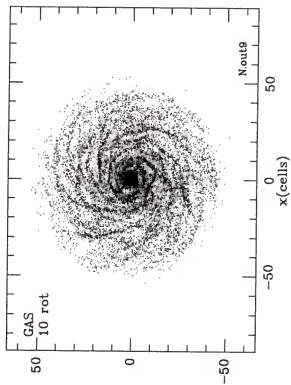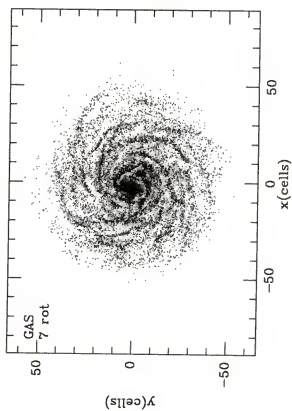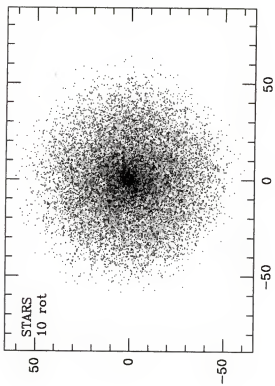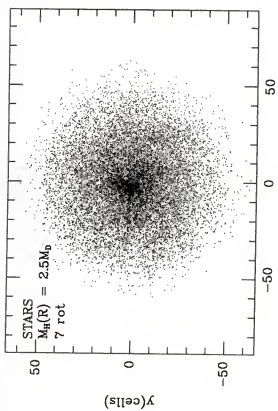
Figure 4-6: The final model in a series of cold two-component models with haloes. Gas makes up 20% of the disk. $\alpha = 0.15$, $\sig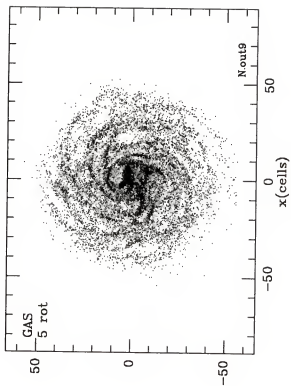ma = 0.6$ cells. The halo is 5 times as massive as the disk. Notice the small-scale structure found in the gas spiral arms.
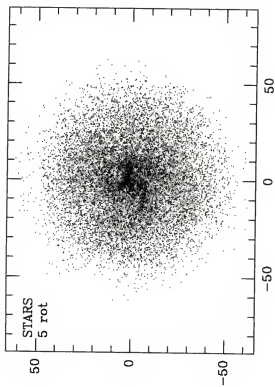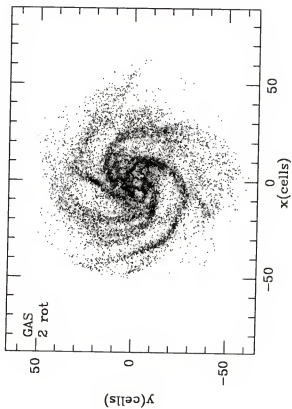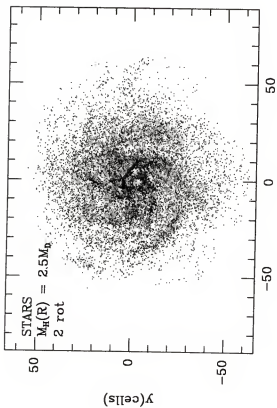
STARS
$M_M(R) = 5M_D$
2 rot

STARS
5 rot

GAS
2 rot

GAS
5 rot

Nout13

x(cells)

y(cells)

Figure 4-6: continued

STARS
$M_H(R) = 5M_D$
7 rot

STARS
10 rot

GAS
7 rot

GAS
10 rot

N.out.13

y(cells)

x(cells)

and allowing the disk to stabilize with smaller noncircular velocities and to respond to shorter and shorter wavelength perturbations. In the gas, the number of spiral arms increases and the arms get narrower, more distinct and longer.

A number of other experiments were run. One series of tests compared responses for the two gas distributions. For completely cold models, similar behavior was expected from both, especially for lower halo mass cases. Because th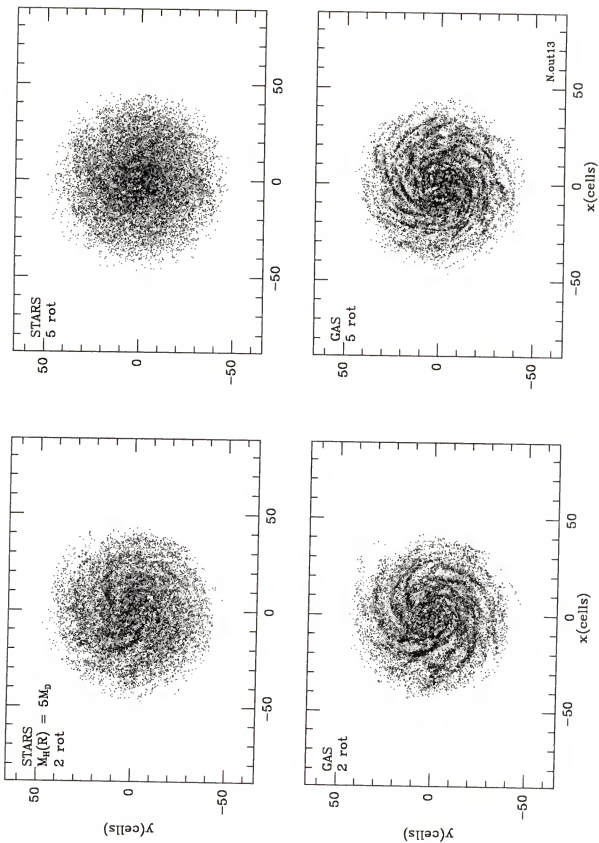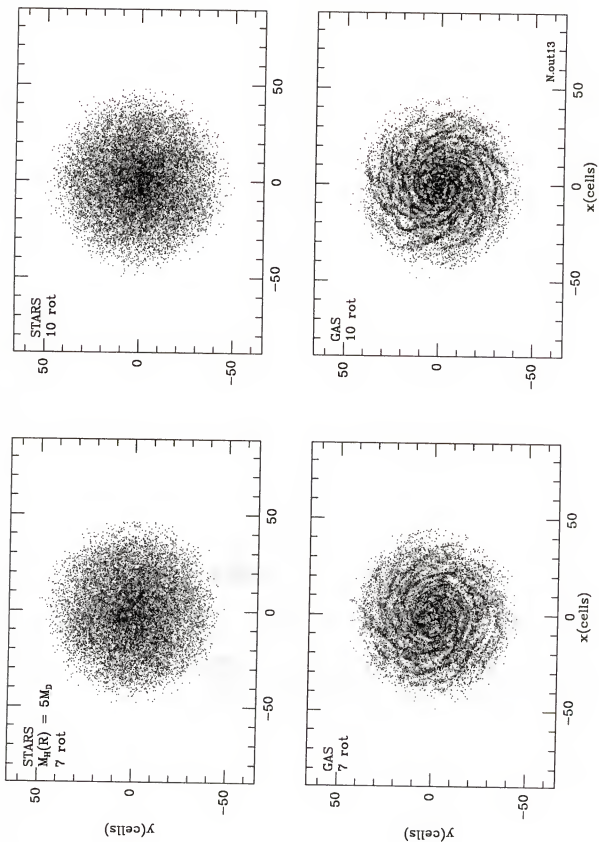e cold disks are violently unstable, the models readjust significantly initially, changing the surface density so much that differences between models are expected to be erased. When density plots of identical models run with the different gas distributions are compared, they look fairly similar. All the conclusions reached for the Toomre gas distribution were found for the constant gas distribution.

A test of the effect of the viscous force amplitude parameter, α, yielded the expected result that as α increased, increasing the magnitude of the gas viscous force, the gas spiral structure became sharper, narrower and more obvious. Figure 4-7 shows the gas distribution after 12 rotations for four values of alpha, ranging from 0.02 through 0.15. In all cases, even for $\alpha = 0.02$, spiral structure is present and it becomes more apparent as the value of α rises. In addition, the gas noncircular velocities decrease as α increases, but even with low values of α, the gas cools itself efficiently. Figure 4-8 is a plot of the gas radial dispersion versus radius after 12 rotations for the lowest and highest values of α shown in figure 4-7. To determine the appropriate α for a model, an attempt was made to select a value that produced gas viscous forces between one hundredth and one tenth of the gravitational forces. This changes from model to model but, roughly

Figure 4-7: The effect of the viscous force amplitude parameter, α, on the gas spiral arms. α varies from 0.02 to 0.15.
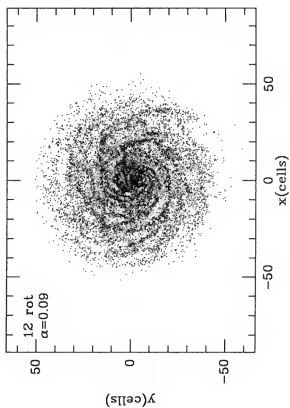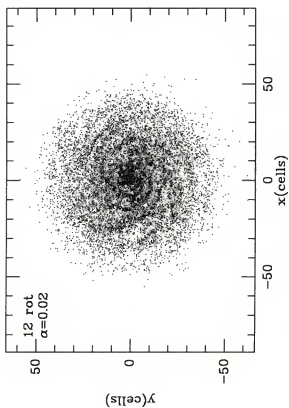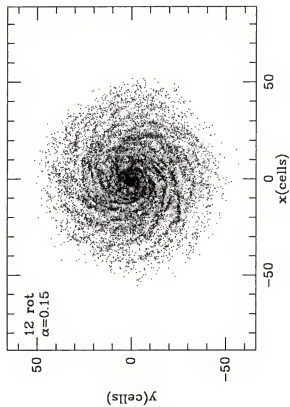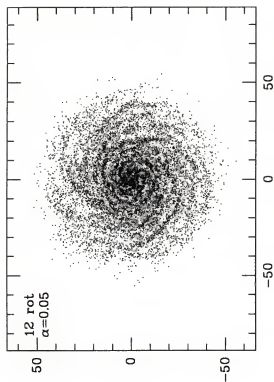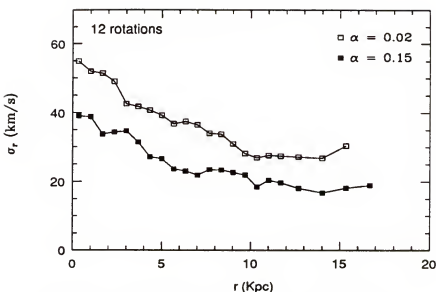
Figure 4-8: The gas radial dispersion for two values of α.

speaking, α was usually less than the 0.15 and greater than 0.02, ranging most often between $0.07 - 0.1$. There are a number of ways to determine a realistic value of α. The first is to compare the arm, interarm contrasts seen in the models with gas observations. This contrast increases as α increases, and should provide a way to constrain α. A second test is to compare the gas peculiar velocities with observed gas dispersions. The gas noncircular velocities decrease as α increases and provide a mechanism for checking α.

When looking at figures for models that contain a gas component, it should be kept in mind that the density and not the luminosity is plotted. As spiral arms in real galaxies contain many hot young stars and HII regions, the luminosity in these regions greatly exceed that of the surrounding areas, even if the density may not.

A final model set calculated the response of increasing the gas mass from 1 to 50 percent. In all models, the gas was initialized in a Toomre distribution, the halo mass

was 2.5 M_D and $\alpha = 0.09$. As the gas mass is increased, the spiral structure gets stronger, self-gravity enhances the spiral response. However, neither the gas nor the stellar velocity dispersions vary much among the models. In addition, increasing the gas mass did not encourage the stellar spiral structure to remain for a longer time. The stellar heating did not decrease and thus stellar spiral activity was not helped by the gas gravitational potential. It appears that the dissipative forces that cool the disk play a more important role in the maintenance of spiral structure than gravitational forces.

Varying Heating Across the Disk

A second method to induce the formation of bars and spiral structure is to reduce the stabilizing stellar peculiar velocities in the central region of the model while maintaining them in the outer disk. A stellar bar then forms naturally in the interior and is surrounded by a stable outer stellar disk. The noncircular velocities in this outer disk can be quite high, preventing the maintenance of any spiral structure that might have developed from the unstable start conditions or from the motion of the bar through the disk. The gas, initialized with little or no velocity dispersions, responding to the stellar potential, also forms a bar in the interior. In the outer gas disk, the viscous forces act to keep the gas cool, thus allowing spiral structure that forms to be preserved. The motion of the bar rotating through the gas disk will act as a continuous source of excitation for creating spiral structure in the dissipative gas medium. The gas component forms spiral features that persist as long as the models are calculated. This procedure to excite a spiral response is self-consistent; no perturbing forces are imposed.

Because isotropic heating can be calculated analytically, it is ideal for these experiments and all cases with a non-constant value of **k** use isotropic heating. It is simple to calculate the heating in the center of the disk, via equation (3-23), corresponding to a high value of **k**, vary it smoothly across the disk so that the outer regions have a lower **k**, corresponding to a larger stabilizing velocity dispersions. When **k** = **k**(r), equation (3-23) is an approximate result, which has been found to work well in practice. If equation (3-22) is integrated, allowing **k** to be a linear function of the radius, **r**, the resulting dispersion equation agrees well with that of the heuristic approach in the outer region of the disk but yields peculiar velocities in excess of those desired in the center of the disk.

The models formed using this procedure were found to be sensitive to a number of parameters. Most important are; the size of the unstable inner region, the slope of the linear function used for **k**, and the halo mass. Many models of this sort have been computed. Three typical calculations have been chosen as representative examples. They will be discussed in turn.

Model 1

In model 1, the initial radius is 36 cells, corresponding to a physical length of 18 Kpc. The total disk mass is 1×10^{11} M$_o$, of which 10% is gas. A halo with a mass equal to the disk mass is imposed. The gas parameters are; $\alpha = 0.1$, $\sigma = 0.7$ cells. The gas is distributed uniformly across the disk and is given a constant dispersional velocity of 5 km/s. The stellar heating varies across the disk from a **k** of 0.98 for **r** less than 12 cells, to **k** = 0.65 for **r** greater than 24 cells. The stellar dispersions vary by roughly a factor of 3 from the inner regions, with values of 25 km/s, to a point midway across the

disk, where the dispersion peak at approximately 65 km/s. The rotation curve of the gas is reduced by a constant **k** value of 0.82. Figure 4-9 shows the initial "rotation curves" for the model. In figure 4-10 the surface density of the stars (top) and gas (bottom) are presented. The number of rotations completed is given on each individual plot.

Gas and Stellar Initial Rotation Curves
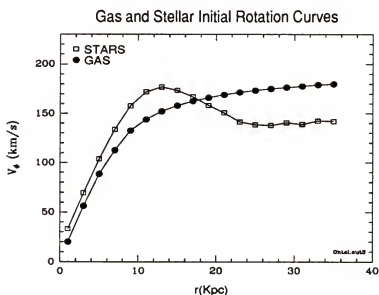
Figure 4-9: Initial rotation curves for the two disk components.

As figure 4-10 shows, early in the evolution of the model, both components develop a hole at the center, with a radius of slightly under 12 cells, the size of the area with reduced heating. However, although this inner region has low noncircular velocities, it does not respond as a cold disk does (Chapter 3, figure 3-1) because as particles in this unstable area start to move outward, they run into the warm outer disk. This pileup of the outward moving material with the outer disk is evident at 0.5 rotations where an over-dense ring is clearly seen. The ring is roughly three cells in width. In this particular

Figure 4-10: Model 1, a two-component model in which the heating varies across the disk. The gas amounts to 10% of the total disk mass and a halo equal in mass to the disk is included. k varies from 0.98 for r less than 12 cells, to $k = 0.65$ for r greater than 24 cells. $\alpha = 0.1$, $\sigma = 0.7$ cells.

STARS
1.5 rot

GAS
1.5 rot

STARS
0.5 rot

GAS
0.5 rot

0ntel.3

x(cells)

x(cells)

y(cells)

y(cells)

Figure 4-10: continued

STARS
5 rot

GAS
5 rot

Ontel.3

STARS
3 rot

GAS
3 rot

y(cells)

x(cells)

Figure 4-10: continued

STARS 7 rot

STARS 9 rot

GAS 7 rot

GAS 9 rot

Ontcl.3

y(cells)

x(cells)

Figure 4-10: continued

STARS
13 rot

STARS
11 rot

GAS
13 rot

GAS
11 rot

0ntcl.3

x(cells)

y(cells)
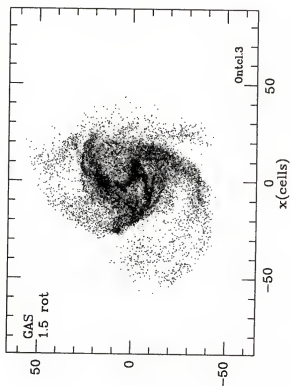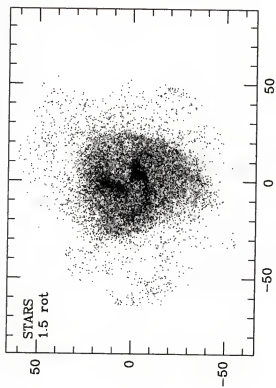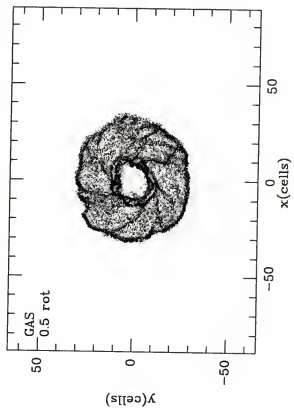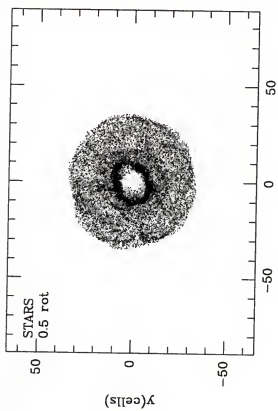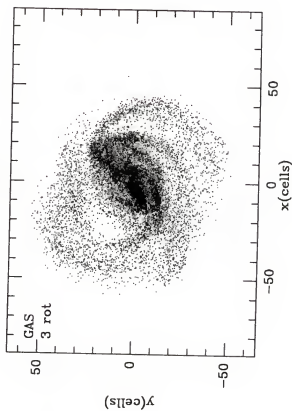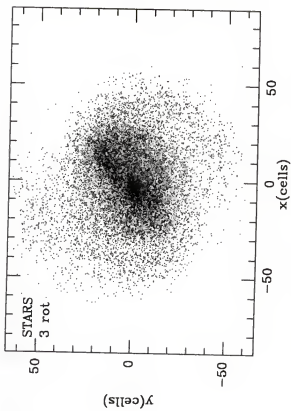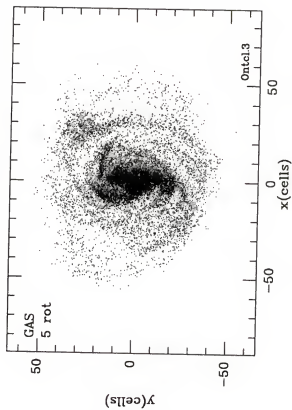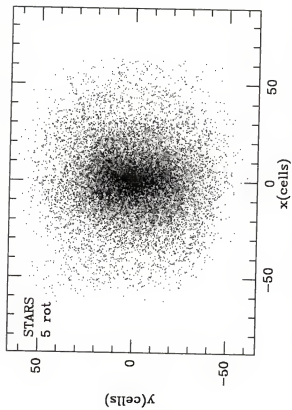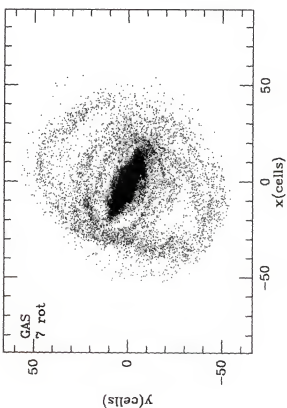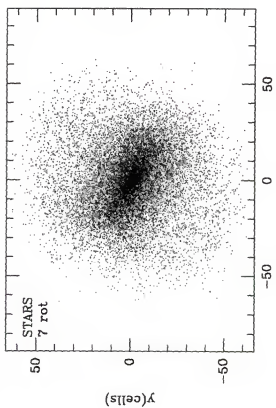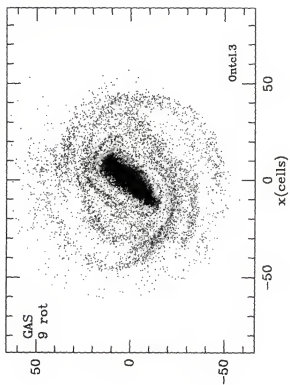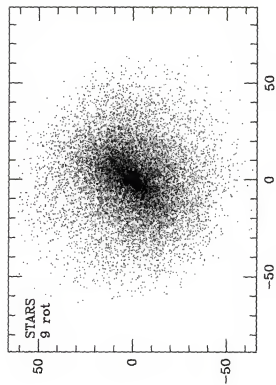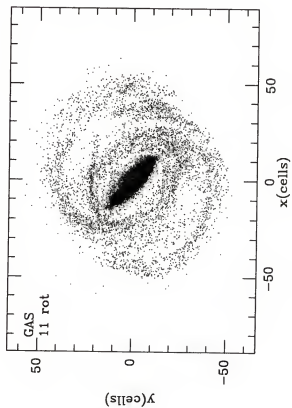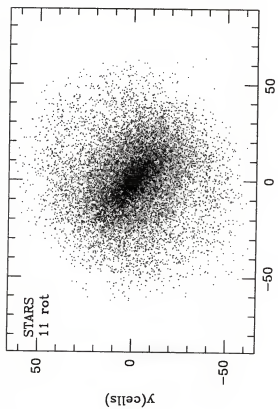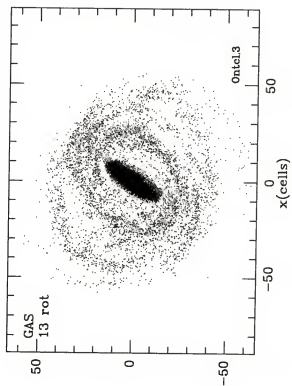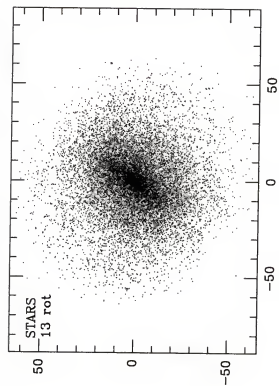
model the initial **k** in the outer regions does not provide enough heating to prevent all nonaxisymmetric developments and the outer stellar disk displays slight asymmetries which become apparent by 1 rotation (not shown). At this point the particles have moved back into the central region in a non-uniform manner; a three armed density enhancement is found. At 1.5 rotations, the stellar disk has begun to spread and become more asymmetric. However, as figure 4-11 shows, the heating has increased significantly by this time, stabilizing the stars. By 5 rotations the stellar development seems to be complete. The dense regions have converged into a long but relatively weak bar that is surrounded by an apparently featureless outer disk. The mass within the bar radius is approximately one-third the total disk mass. Most of the asymmetry in the outer disk has disappeared and there is little change in the stellar component after this time.

In the gas component, initialized with little heating, and distributed uniformly, the early development is similar to that of the stellar component. At 0.5 rotations, the gas disk
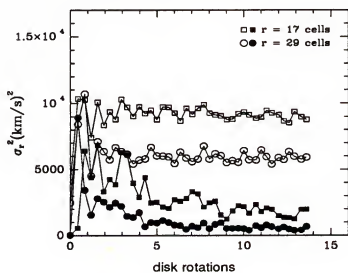
Figure 4-11: Gas (filled symbols) and stellar (open symbols) peculiar radial velocity squared for two radii in the outer disk.

has a severely depleted central region, surrounded by a ring of enhanced density. Due to the initial uniform gas distribution, which makes the central gas density considerably less than that of the stars, the gas ring is not as dense as its stellar counterpart. In the outer disk however, due to the lack of heating, the gas has responded to small-scale perturbations, similar to those found in cold disks. At 1 rotation, (not shown) the gas response looks much like the stellar. By 1.5, the effects of the viscous force are becoming noticeable; the "arm-like" features are much more clearly defined in the gas component. In subsequent development, the gas tends to follow the global stellar distribution but the dissipative forces allow the gas to responds much more strongly to small density perturbations in the outer disk. By 3 rotations, spiral features are found. These start at the ends of the bar and continue to become more defined as the model progresses. By 7 rotations, a ring feature has emerged, probably associated with the outer Lindblad resonance. This feature is seen for the rest of the calculation, through 13 rotation periods.

The effect of the bar on the radial velocity is apparent in figure 4-12. In the inner regions (r less than 10 Kpc, 20 cells) the noncircular velocities are very high, but drop off substantially in the outer disk, especially in the gas component. From figure 4-10, the gas bar ends at roughly 20 cells. It contains just under 40% of the disk mass.

Figure 4-13 shows the rotation curves after 12 rotations for the gas and stars. The number of particles in the outer regions is low, making the velocities in these areas suspect, with estimated errors between 20 to 30km/s. The gas rotation curve can be believed out to approximately 8 Kpc, and the stellar, perhaps to 10. However, even with the large errors, in all three models presented, the final gas rotation curve is higher

than the stellar in the outer disk. This could be due to the way the rotation curves are initialized, however, a limited number of models that were calculated with the two components having identical initial rotation curves also ended up with higher gas rotation curves.
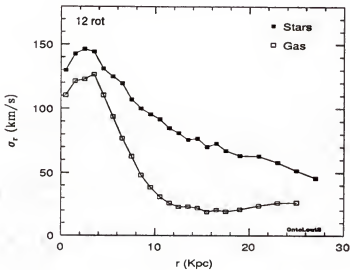
Figure 4-12: The gas and stellar radial dispersion, after 12 rotations.
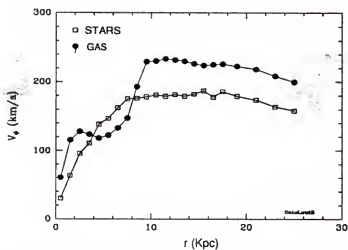
Figure 4-13: Gas and Stellar rotation curves after 12 rotations

<u>Model 2</u>

In the second model a more massive halo is imposed and the initial stellar disk is given more stabilizing peculiar velocities than model 1. The initial radius is 36 cells, 18 Kpc, and the total disk mass is 1×10^{11} M_0 of which 10% is due to the gas. A halo of mass 1.6 $M_D(R)$ is imposed. The value of r_0 is 0.2 R, and r_e is 1.8 R. The gas parameters are; $\alpha = 0.1$, $\sigma = 0.7$ cells and the gas component has a constant dispersion of 5 km/s. The disk contribution to the gas rotation curve is reduced by a **k** of 0.78. The stellar heating varies across the disk, from a **k** of 0.98 for r less than 5 cells to 0.55 for **r** greater than 28 cells. Density plots of this model are seen in figure 4-14.

The model development is similar to that of model 1 but on a less violent scale. At 0.5 rotations, the size of the hole in the center is smaller, on order of 5 cells. Because of this, the gas disk is more extensive, allowing it more area to display sharp spiral like features, caused by differential rotation. At 1.5 rotations, the disk is more symmetric, in both inner and outer areas, than model 1 at the same time frame. It also shows less density enhancement in the center. By 3 and 5 rotations, spiral structure appears in the gas component but it is not as open as was that in model 1. Rather, the disk has spread out very little, forming dense spirals fairly close to the model center. The inter-arm region is densely populated. By 5 rotations a small bar is beginning to appear. Between 5 and 9 rotations, the model expands slightly, and the ring feature found at three rotations begins to differentiate into narrow spirals. These remain close to the bar however, rather than displaying the open structure of model 1. The lower mass halo combined with the smaller stabilizing noncircular velocities of model 1 lead to a more open spiral structure.

Figure 4-14: Model 2, a two-component model in which the heating varies across the disk. The gas amounts to 10% of the total disk mass and a halo 1.6 times as massive as the disk is included. k varies from 0.98 for r less than 5 cells, to $k = 0.55$ for r greater than 28 cells. $\alpha = 0.1$, $\sigma = 0.7$ cells.

STARS
1.5 rot

GAS
1.5 rot

STARS
0.5 rot

GAS
0.5 rot

Ontcl.5

y(cells)

x(cells)

x(cells)

Figure 4-14: continued

Figure 4-14: continued

Figure 4-14: continued

STARS
11 rot

STARS
12.5 rot

GAS
11 rot

GAS
12.5 rot

Ontel.5

y(cells)

x(cells)
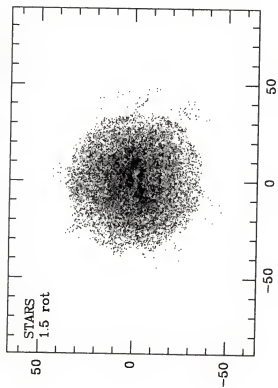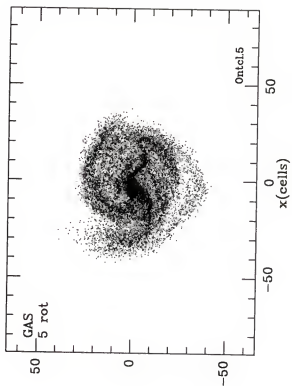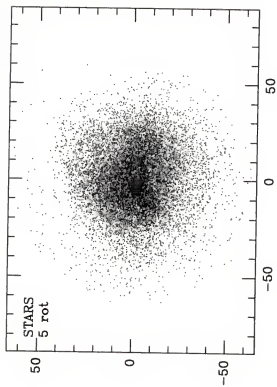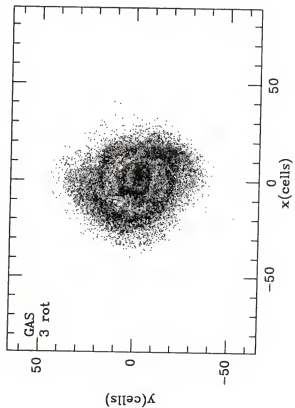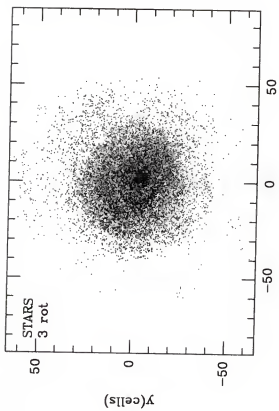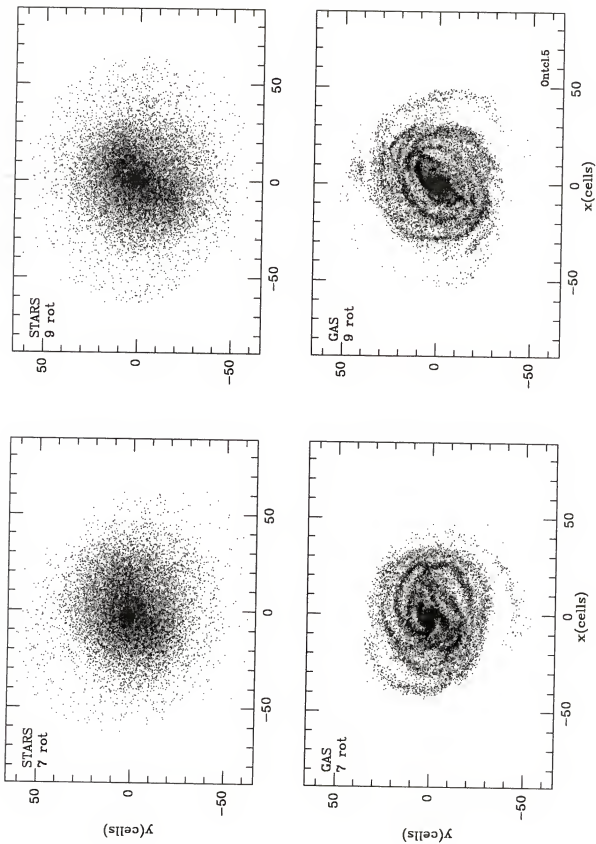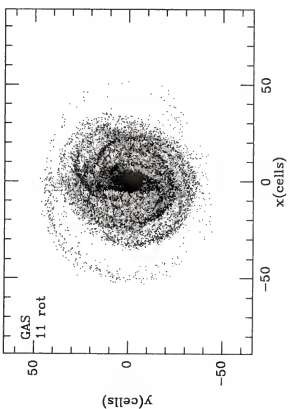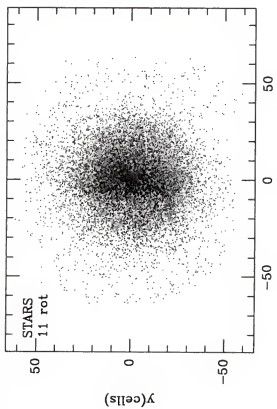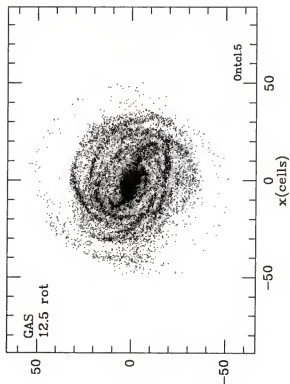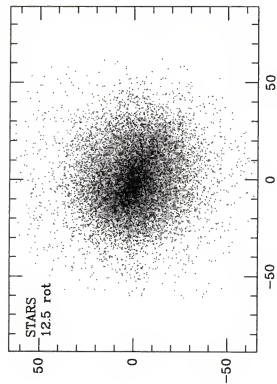
In addition, the spirals in model 2 are narrower, more sharply defined and contain more particles than those in model 1. The bar appears to be smaller, in keeping with the idea that this represents a more "stable" model than model 1. Approximately one-third of the disk mass is contained within an area of radius 20 cells. The gas and stellar peculiar radial velocity after 12 rotations are seen on the left in figure 4-15 and the rotation curves at the same time, on the right. Again, the values of the rotation curve beyond 8 to 10 Kpc, are highly suspect. The stellar radial peculiar velocities are higher than observed values but by less than a factor of 2 over the outer half of the disk.
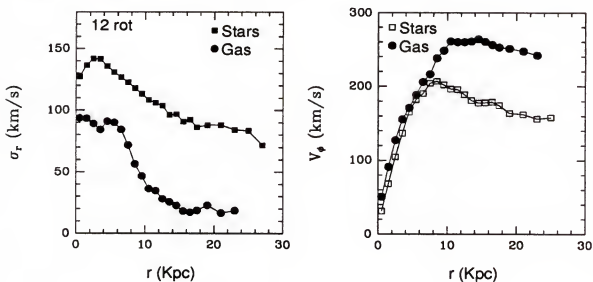
Figure 4-15: The radial peculiar velocity and
rotation curves for both components after 12 rotations.

Model 3

As was the case in models 1 and 2, the initial radius of model 3 is 36 cells, (18 Kpc), with a total disk mass of 1×10^{11} M$_\odot$. 5% of this is due to gas and a light halo of mass 0.6 M$_D$(R), with a r_o and r_e of 0.2 R and 1.8 R, is imposed. The gas parameters

are; $\alpha = 0.1$, $\sigma = 0.7$ cells. The gas component is put in with a constant dispersion of 5 km/s, and the disk contribution to the gas rotation curve is reduced by a **k** of 0.78. The stellar heating varies across the disk, from a **k** of 0.98 for **r** less than 4 cells to 0.5 for **r** greater than 24 cells. Density plots of this model are seen in figure 4-16.

Although the inner cold region is initialized in a way similar to that of model 2, the outer regions have more heating and so early in the model development, the disk is relatively stable despite the low halo mass. At half a rotation period, there is no hole blown out of the central regions; instead the density is slightly enhanced. By 1.5 rotations, the model has begun to expand outward. From 3 to 5 rotations, the development of the bar is seen. During this time, the outer regions continue to spread outward while the inner regions collapse into a long narrow bar. At 5 rotations a lenticular shape appears which persist through the calculation. An inner and outer set of arms (rings) are found. The inner spirals start at the ends of the bar and "hammerhead" features are seen at points in the model development. The outer spiral features are very tenuous, due to the fact that many of the gas particles have been trapped in the bar or lost off the edge of the grid. In agreement with the discussion on model 2, it appears that more open spiral features are found when a less massive halo is imposed. The rotation curve for model 3 after 12 rotation curves is shown in figure 4-17.

An attempt was made to isolate the effects of the halo mass in the varying heating models. To test the hypothesis that as the halo mass increased the spiral structure got less open (and more defined, probably due to the greater particle density), a series of models were calculated allowing the halo mass to increase from 0.6 to 2.5 M_D as all

Figure 4-16: Model 3, a two-component model in which the heating varies across the disk. The gas amounts to 5% of the total disk mass and a halo 0.6 times the disk mass is included. k varies from 0.98 for r less than 4 cells, to k = 0.5 for r greater than 24 cells. α = 0.1, σ = 0.7 cells.
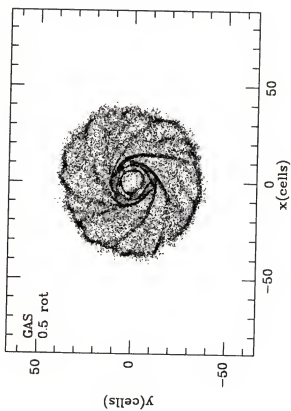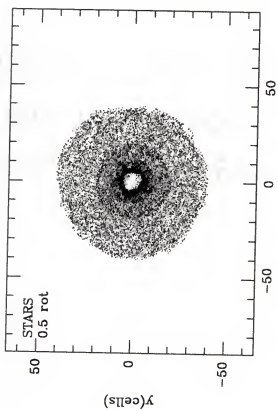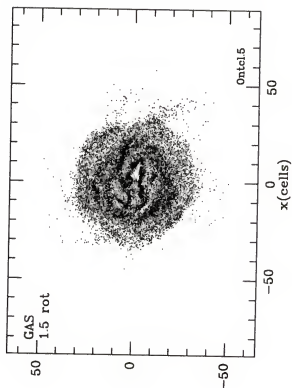
STARS
1.5 rot

GAS
1.5 rot

nN.out7

STARS
0.5 rot

GAS
0.5 rot

x(cells)

y(cells)
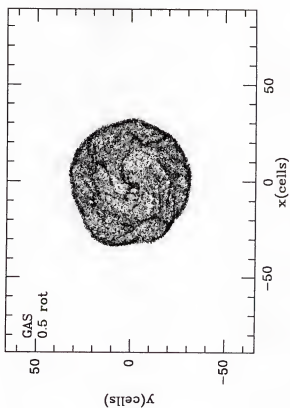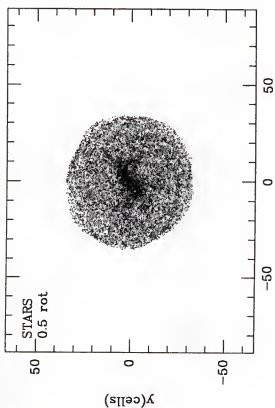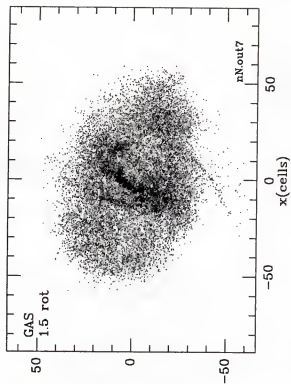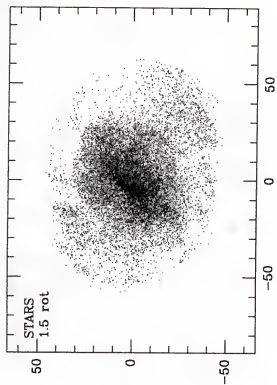
Figure 4-16: continued
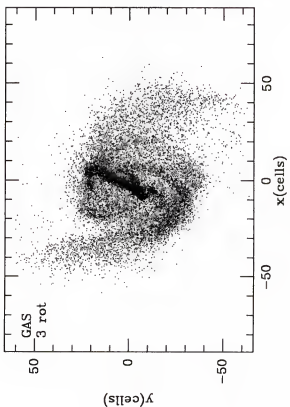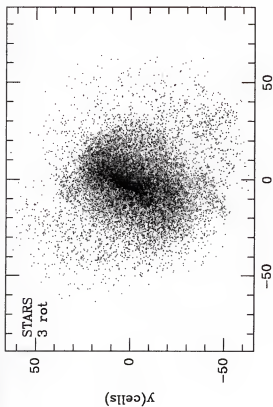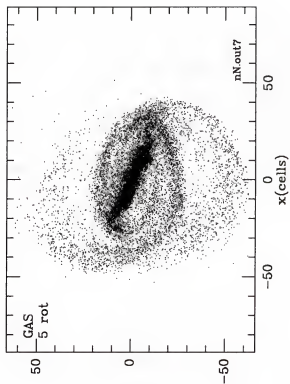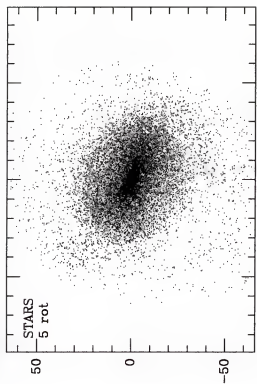
Figure 4-16: continued
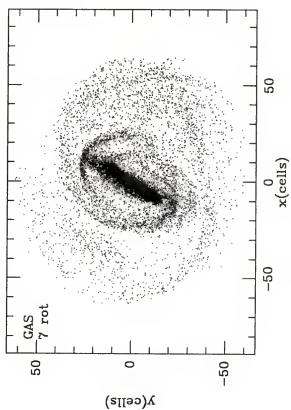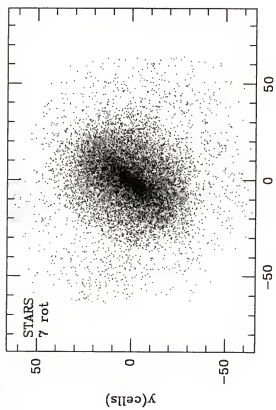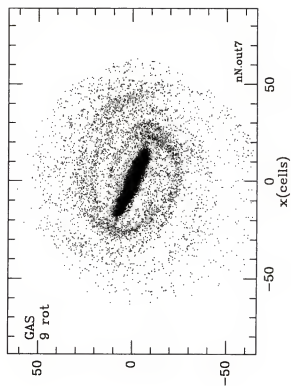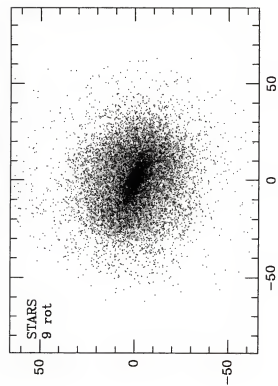
Figure 4-16: continued
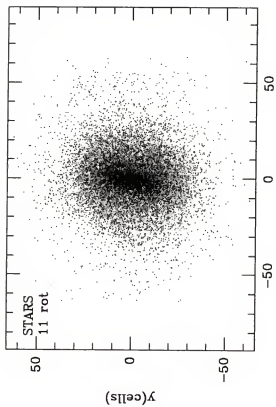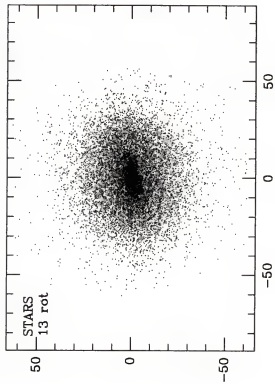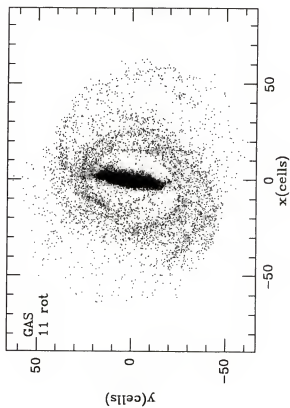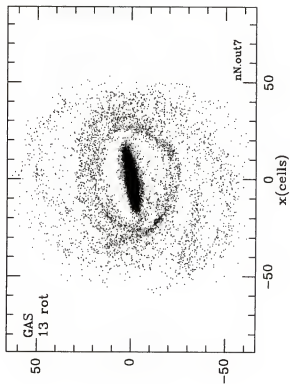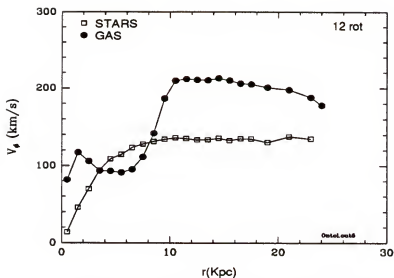
Figure 4-17: Gas and stellar rotation curves after 12 rotations.

other parameters were kept constant. Figure 4-18 shows the gas component of two of these models at 10 rotations. The arms appear to become more open as the halo mass decreases.

Figure 4-19 shows the gas component after 12 rotations for two models that are identical except for the value of the softening parameter. On the top, $\epsilon = 0.5$, on the bottom, 3.0. The larger softening length figure shows less open structure and less of a bar than the lower. This trend continues when the softening length is increased to 4 cells (not shown). As the softening length increases, the model is more stable, the bar is less pronounced, and the arms closer in to the bar.

Finally, an effort was made to find evidence of stellar spiral structure in contour plots of the gravitational potential, which was dominated by the stellar component in most cases of interest. For the most part, these efforts were unsuccessful. No spiral structure was evident in the potential beyond one or two rotation periods. A second attempt to
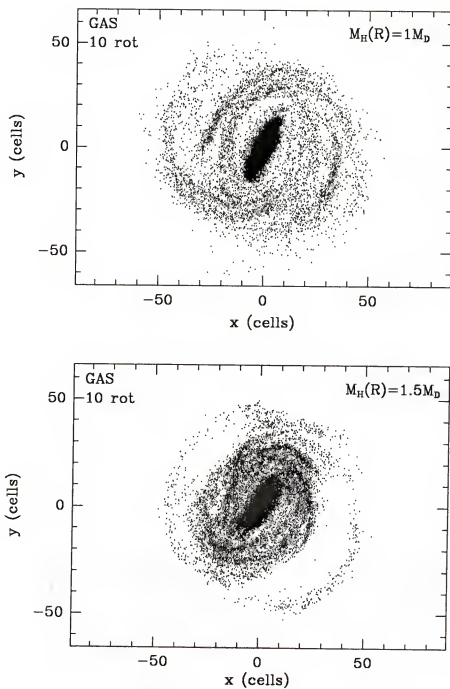
Figure 4-18: The effect of the halo mass on the gas response of a two-component model in which the heating varies across the disk.

find stellar spirals was to make density plots of only the stars which had low dispersional velocities. Although the spiral structure was more evident on these maps than on plots including all the stars, in no models were the stellar spirals very strong or obvious.

In this chapter two methods of developing models with bar and spiral features were presented. The second method, a new technique, involves destabilizing the stellar disk in the central regions to create a bar. The bar moving through the background disk should destabilize the disk to spiral perturbations. In the stellar component, the heating is so large, that any spiral structure is quickly damped out. In the dissipative gas component, however, the peculiar velocities are low enough to allow spiral structure to persist. The resulting models are sensitive to the size of the cold region in the center, to the rate at which the heating is "turned on" over the disk, and the halo mass. In most models all these factors compete, giving a wide variety of spiral structure. If the halo is light, or the heating not very great, open spiral features are found. Conversely, if a massive halo is added, very closed spiral arms develop. If too much heating is applied, the model tends to be featureless.

This method of inducing structure demands that a cooler interior exist in the disk. If a proto-galaxy is made up of cold molecular gas that cools itself efficiently in the dense central regions, this is not an unreasonable start condition.
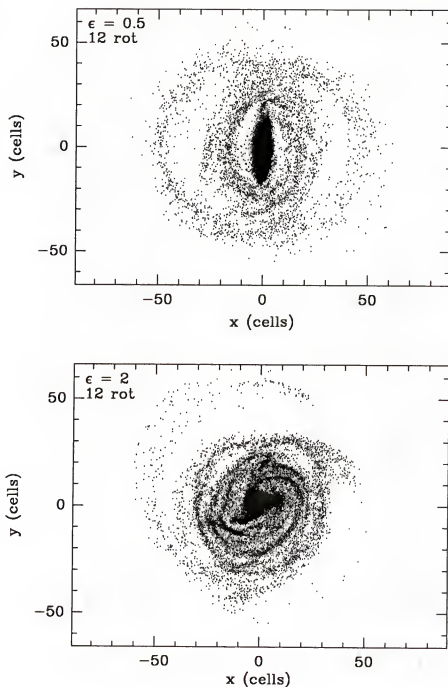
Figure 4-19: The effect of ϵ on the gas response of a two-component model in which the heating varies across the disk. A halo equal in mass to the disk mass is included.

CHAPTER 5
NEUTRAL HYDROGEN OBSERVATIONS

One of the incentives for developing the code described in the last three chapters is
to use it to model particular barred spiral galaxies. It is hoped that observations can be
used to constrain model parameters. The stellar bar can be compared to infrared data
while the projected radial velocities of the gas component can be compared to 21 cm
VLA radio observations. In this chapter, the observation and data reduction of 21 cm
HI data of 2 barred spiral galaxies, NCG 1398 and NGC 1784 will be discussed.

Neutral Hydrogen as a Kinematic Tracer

For spiral galaxies, the gas component, although a minor component by mass
fraction, is important from an observational standpoint. Because the gas responds so
nonlinearly to small perturbations, it is believed that the gas distribution and kinematics
are sensitive tracers of the underlying gravitational potential (Roberts, Huntley and van
Albada 1979). Observations of the gas should provide information on the overall galactic
potential and thus are useful when comparing numerical models to actual galaxies; a
successful model should match the rotation curve across the disk, derived from the gas,
and the gas surface density distribution of the real galaxy.

There are several possible tracers of the gas in a spiral galaxy. Ideally a tracer
would be distributed throughout the entire galaxy and would provide global kinematic
information. Unfortunately, there is no one gas component of the interstellar medium

187

(ISM) that lives up to this ideal. The largest component of gas by mass in the ISM of a spiral galaxy is in the form of cold, (T = 10 - 100K) dense, molecular clouds. Made up predominantly of molecular hydrogen and a significant amount of carbon monoxide, these clouds, unfortunately, are concentrated in a small fraction of interstellar space. A second important component of the ISM is hot, ionized hydrogen gas, (HII) surrounding groups of hot, young stars. HII regions make up approximately ten percent of the ISM by volume, mainly in spiral arms or near the galactic center. Finally, neutral hydrogen (HI) envelops the cooler interstellar clouds and HII regions and fills about 20% of interstellar space.

Not many of these components are ideal for observing. Molecular hydrogen is extremely difficult to detect. Molecular carbon monoxide (CO) emits line radiation in the millimeter regime and has been used to mark giant molecular clouds within our galaxy. These observations show that the CO is not distributed uniformly but appears to be concentrated in a ring of approximately 3 Kpc width, at a mean radius of 5.5 Kpc (Mihalas and Binney, 1981). There is a sharp drop in density on either side of this regions. In the very inner regions (r < 300 pc), the molecular cloud density is again high. This nonuniform distribution is a drawback to using CO as a kinematic tracer over an entire galactic disk. In addition, CO maps of other galaxies are difficult to make, as the sensitivity of available millimeter telescopes is still low. A third candidate for observations is the optical alpha line of ionized hydrogen. Such observations can be of very high resolution, which is not necessary considering the lack of sophistication of the models, but are incomplete due to the patchy nature of these regions.

The final and best option is neutral hydrogen. It permeates galactic disks and is one of the flattest and thinnest components, especially in the inner regions, of both our own and other galaxies. There is a hyperfine transition in the ground state of neutral hydrogen that results in a 21cm radio spectral line. This line is relatively easy to detect, and does not suffer from either interstellar or self absorption in general. In our galaxy, the distribution of HI is close to flat between 5 and 13 Kpc. In other galaxies it extends out further than HII or CO regions and falls off in density more slowly than either stars or HII regions. It therefore can be used to provide maps of the integrated column density of the HI gas and the distribution of mean radial velocity across the disk.

Much of the interstellar medium consists of neutral hydrogen atoms in the ground state. This state is split into two levels, depending on the orientation of the electron spin vector with respect to that of the proton. The energy difference between the levels is small, the energy for parallel spins is 6×10^{-6} eV (corresponding to a temperature difference of 0.07K) higher than that of antiparallel spins. When an electron in the higher state makes a spontaneous "spin flip transition" to the lower energy level, a 21.105 cm (1420.4 MHz) radio photon is emitted. As the temperature required to put the electron in the higher state is small, most of the hydrogen atoms are in the excited state. Because the spin flip transition is forbidden, it occurs rarely, only once every 10^7 years. The collisional de-excitation rate is much more rapid, once every 400 years (for $N = 20$atoms/cm^3). Thus collisions establish equilibrium populations in the two levels which can be described using the Boltzmann formula:

$$\frac{n_2}{n_1} = \left(\frac{g_2}{g_1}\right) \exp\left(\frac{-h\nu}{kT_s}\right) \tag{5.1}$$

where g_i is the degeneracy of the level and T_s is the spin temperature of the gas. In a typical HI region, T_s is close to 100K, which yields a small value of the argument for the exponential and a value of approximately one for the exponential itself.

It is possible to derive a simple relation between the observed brightness temperature and the column density of neutral hydrogen using the radiative transfer equation. Following Mihalas and Binney (1981), the transfer equation through an element of length ds can be written as:

$$\frac{\partial I_\nu}{\partial s} = \frac{h\nu}{4\pi}[n_2 A_{21}\phi_\nu - (n_1 B_{12} - n_2 B_{21})\phi_\nu I_\nu]$$
$$= \eta_\nu - \chi_\nu I_\nu \tag{5.2}$$

where

n_i represents the number of atoms in the ith level.

A_{21} is the Einstein probability coefficient for spontaneous radiative decay.

B_{21} is the Einstein probability coefficient for induced decay.

B_{12} is the Einstein probability coefficient for absorption.

χ_ν is the absorption coefficient.

η_ν is the emissivity of the material.

ϕ_ν is the fraction of atoms in either level that can absorb or emit radiation of frequency ν. $\int \phi_\nu d\nu = 1$.

The Planck function:

$$B_\nu = \frac{\eta_\nu}{\chi_\nu} \tag{5.3}$$

and the optical depth:

$$\tau_\nu \equiv -\int \chi_\nu ds. \tag{5.4}$$

can be substituted into the transfer equation simplifying equation (5.2) to:

$$\frac{\partial I_\nu}{\partial \tau_\nu} = I_\nu - B_\nu. \tag{5.5}$$

Solving this equation yields:

$$I_\nu = B_\nu[1 - \exp(-\tau_\nu)] + I_{\nu o}(\tau_\nu)\exp(-\tau_\nu) \tag{5.6}$$

where $I_{\nu o}$ is the incident intensity on the far side of the material and is assumed to be zero. In the radio regime, the Rayleigh–Jeans approximation can be used to write the Planck function as $B_\nu(T_s) \approx \frac{2kT_s}{\lambda^2}$. The brightness temperature, T_B, is defined in terms of the observed intensity, $T_B \equiv \left(\frac{\lambda^2}{2k}\right)I_\nu$. Substituting these into equation (5.6) simplifies it to give:

$$T_B = T_s[1 - \exp(-\tau_\nu)]. \tag{5.7}$$

From equations (5.1), (5.2) and (5.4), the optical depth in the 21 cm line can be written as,

$$\tau_\nu = \frac{CN_H\phi_\nu}{T_s} \tag{5.8}$$

where the constant C has the value 2.57×10^{-15} cm^2 K^{-1} and

$$N_H = 4N_1 = 4\int_0^\infty n_1(s)ds \tag{5.9}$$

where N_1 is the integrated column density of hydrogen in the lower level along the line of sight. To get the column density (atoms cm^{-2}), of the gas at a point, equation (5.8)

is integrated over all frequencies (velocities) and solved for N_H:

$$N_H = 1.82 \times 10^{18} \int\limits_{0}^{\infty} T_s \tau_\nu dv. \tag{5.10}$$

The units of the velocity are km/s. From equation (5.7), for the optically thin case, τ_ν $\ll 1$, and $T_s \tau_\nu$ can be replaced by the observed brightness temperature, T_B (Mihalas and Binney, 1981, p. 489).

$$N_H = 1.82 \times 10^{18} \int\limits_{0}^{\infty} T_B dV. \tag{5.11}$$

The approximation of an optical thin gas was acceptable for both of the galaxies observed. The highest brightness temperatures, averaged over the beam, were 27.0 K for NGC 1398 and 21.3 K for NGC 1784. These yield maximum optical depths of 0.32 and 0.24 respectively, assuming a spin temperature of 100K.

Thus observations of the 21 cm line intensities give the column density along the line of sight. To determine the total mass of neutral hydrogen in the galaxy, the column density is integrated over the area of the galaxy.

In addition to the neutral hydrogen distribution, Doppler shifts in the 21 cm observations yield the rotation curves of external galaxies. For a galaxy inclined at angle i to the line of sight, the observed radial velocity will be

$$v_R(\rho, \phi) = v_o + \Pi(R, \theta) \sin \theta \sin i + \Theta(R, \theta) \cos \theta \sin i + Z(R, \theta) \cos i \tag{5.12}$$

(Mihalas and Binney, 1981, p. 499) where (R, θ) are in the plane of the galaxy, (ρ, ϕ) are in the plane of the sky, v_o is the systemic velocity, Π and Θ are the radial and tangential

velocities in the plane of the galaxy and Z is the velocity component perpendicular to the plane. This is solved for $\Theta(R,\theta)$.

The mean temperature weighted velocity at a point, is given by the first moment of the brightness temperature with respect to velocity: (England, 1986).

$$\langle V(x,y) \rangle = \frac{\int\limits_{\infty}^{\infty} T_B(x,y)V(x,y)dV}{\int\limits_{\infty}^{\infty} T_B(x,y)dV} \qquad (5.13)$$

Program Galaxies

The galaxies studied at the VLA were NGC 1398 and NGC 1784. Apart from the obvious requirement of being observable at the VLA, the galaxies had to satisfy several criteria concerning size, signal strength, etc. In order to be well resolved, only galaxies larger than 4′ were considered. To be certain of a large signal to noise ratio, past work done at the University of Florida suggested that only galaxies with a HI surface brightness greater than that of NGC 1300 be considered. Isolated galaxies with a prominent bar, large in comparison to the synthesized beam, were sought. Finally, optical or near infrared surface photometry had to be available to constrain the mass distributions of the calculated models.

Brief descriptions of two barred spiral galaxies that fit all the criteria, NGC 1398 and NGC 1784, are given below.

NGC 1784: ($\alpha = 05\ 03.1$, $\delta = -11\ 56.4$). This is a SBc(rs) I-II galaxy with a photometric diameter (measured at a brightness of 25th magnitude per square arcsec) of

4.2′. It has a bright, narrow bar and a ring with dimensions 1.3′ × 0.7′. Further out, there are a number of faint, knotty arms, with many spurs and feathers. The HI surface brightness is quite strong, 55% greater than that of NGC 1300. Surface photometry observations had previously been made by Dr. S. P. Grosbol. NGC 1784 can be seen in the top of figure 5-1.

NGC 1398: (α = 03 36.8, δ = -26 29.9). This is a SBab(r) I (Sandage and Tammann, 1981) galaxy. It is a strikingly symmetric galaxy with a photometric diameter of 6.6′. The bar dimensions are 1.3′ × 0.2′. The galaxy has well developed, thin, tightly wound spiral arms and an outer pseudo ring of dimensions 4.8′ × 3.3′. The HI surface brightness is 18% greater than that of NGC 1300. Dr. R. Buta has made infrared surface photometry observations of this galaxy. NGC 1398 is shown in the bottom of figure 5-1.

The galaxies were observed at 21 cm at the VLA in February 1991 in the C/D configuration and in February 1992 in the B/C configuration. The VLA is an earth-rotation aperture synthesis radio array, made up of 27 steerable dishes, each 25 meters in size. The telescopes are disposed among three arms, spaced 120° apart, orientated approximately in the north, southeast and southwest directions. The spacing of the dishes along an arm increases outwardly, in a power law design, giving a denser sampling at shorter baselines, resulting in better sensitivity. The antennas can be set in four basic configurations, the A, B, C or D, based on the dish spacings. The length of an array arm decreases by a factor of approximately 3 for each configuration. The A array is the largest, with arms up to 20 km long, but it lacks the shorter baselines, the closest dishes

Figure 5-1: NGC 1784 (top) and NGC 1398 (bottom)

NGC 1784

120"

SBbc(rs) I-II

V = 2254

lie 0.5 km from the center. The D array is the smallest, with a maximum arm length of 0.6 km. The placement of the antennas along the array arms determines the resolution and brightness sensitivity of the configuration. A compromise must be reached between these two factors. For an interferometer, higher resolution is obtained by increasing the spacing between antenna pairs. The better the resolution, however, the poorer the brightness sensitivity to extended structure. When used as a spectrometer, the signal is divided into many channels, further lowering the sensitivity. The relatively weak signals expected from the galaxies dictated that observations be made when the array was in the two smallest, lowest resolution configurations, the D and C. The D configuration contains spacings from a maximum of about 1 km down to 40 m. The separations in the C array vary from roughly 100 m up to 3.5 km.

Due to the low declination of both objects, hybrid configurations, C/D and B/C were used to give better u,v coverage and a more circular beam. A hybrid configuration with a long north arm is useful for imaging low declination galaxies. For such galaxies, the north south baselines are severely foreshortened by projection, leaving a blank region about the v axis of the (u,v) coverage. The hybrid configurations fill in this empty region. The C/D hybrid configuration has the north arm extended in the larger C position. Its maximum unprojected baseline is 2107 m and the minimum is 45 m. The B/C is a similar, higher resolution shape with the north arm in the longer B arm. The maximum and minimum baselines are 6920 m and 78 m.

Observations at the C/D configuration were taken for ten hours, split roughly between the two galaxies. The time on object was approximately 4 hours for NGC 1398 and 3.75

for NGC 1784. B/C observations included 8 hours for NGC 1398 and 10 hours for NGC 1784, of which 6.5 and 7.3 hours respectively, were spent on source. A total bandwidth of 3.125 MHz, divided into 32 channels, was used. 31 of these were narrow line channels with a frequency separation of 97.7 kHz (20.6 km/s) per channel. The number of channels must be a power of two (to allow for the use of Fast Fourier Transforms in map making). The value is selected by considering sensitivity requirements, the velocity resolution desired and the velocity range of the global profile of the galaxy under study. The thirty-second channel, labelled channel zero, is a pseudo-continuum channel used to calibrate the line channels. For NGC 1784, the band was centered at 1409.4MHz, corresponding to a heliocentric radial velocity of 2308 km s^{-1}. For NGC 1398, the band was centered at 1411.562MHz, or 1409 km s^{-1}. 2 IFs were used.

The data was edited and calibrated using the NRAO Astronomical Image Processing System (AIPS). The standard reduction outline for spectral line data written by D. Puche and A. Cox of NRAO was followed.

Aperture Synthesis

To achieve high resolution, a telescope receiver must be many times greater than the wavelength observed, ($\theta = \lambda$/D rad), where D is the telescope diameter. For a 21 cm wavelength, the telescope size necessary to get arc minute resolution is prohibitive for a single dish. Thus in the radio regime, an interferometer approach is required if high resolution is desired. The simplest radio interferometer is made up of two separate antennas whose combined signal goes to the receiver. As the Earth rotates, a source

will move relative to the antenna baseline and the combined wavefronts from the two antennas will alternately come in and out of phase due to interference. This gives rise to a series of maxima and minima, fringes, in the received signal, of width $\theta = \lambda/B$, where B is the baseline length between the antennas (Hey, 1983).

Information on the source structure and intensity is obtained by measuring the fringe visibility at many different spacings of the interferometer pair. In terms of Fourier analysis, each interferometer pair measures a single Fourier component of the angular distribution of the source. Combining these measurements, enables one to derive the structure of the source with as much detail as a single antenna of the same aperture dimensions as the widest spacing used in the interferometer would give. This procedure of combining radio signals from different antenna to mimic the effect of a single, much larger dish is called aperture synthesis. It is difficult in practice because to correlate the data from each section requires that the phase relationship between the waves be known. The best way of including this information is to obtain the output of two sections together, i.e., with an interferometer (Hey, 1983).

There are numerous ways to lay out a multiple aerial system in order to simulate the resolution of a large aperture. In order to exactly achieve this, all the relative spacings found in the large aperture must be included. If this is done, combining the signals will yield the same resolution as a single large aperture. Because the total sum area of all the small dishes is much less than the area of the single, larger aperture, only the resolution, and not the sensitivity of the single radio telescope is obtained. The more radio telescopes available to observe at different spacings simultaneously, the

shorter the total observing time required. At the VLA, 27 radio telescopes provide 351 interferometers simultaneously.

Calibration

Before discussing calibration the (u,v) coordinate system used extensively at the VLA will be described. For an antenna pair separated by baseline B, the (u,v) coordinate system is used to describe the projected baselines i.e., the baseline as they would appear for an observer situated at the source. The v axis is parallel to the Earth's rotation axis, and u is perpendicular to this. As the earth rotates the baseline length changes in this system. The projected baseline has an east-west component, u, and a north-south component, v, given in units of wavelength by:

$$u(H, \delta) = B_x \sin(H) + B_y \cos(H)$$
$$v(H, \delta) = -\sin \delta(B_x \cos H - B_y \sin H) + B_z \cos \delta$$

(5.14)

where H is the hour angle and $B_{x,y,z}$ are the x, y and z component of the baseline. The x, y and z axes point towards (H = 0, δ = 0) for x, (H = –6h, δ = 0) for y, and (δ = 90°) for z (Perley, et al, 1989).

For an interferometer, the recorded data comes in the form of observed visibilities, V′. The original signals are collected from the antenna, amplified, converted, transported, correlated, averaged and then recorded; the result is the observed visibility data. The true complex visibility, V, represents the correlated signal response from one antenna pair for one time record. The two may differ for a number of reasons; interference, weather, faulty tracking or antenna equipment etc.. For an extended source, the true

visibilities may be written as:

$$V_{ij}(u,v,t) = \int\limits_{-\infty}^{+\infty} \int\limits_{-\infty}^{+\infty} A_\nu(l,m) I_\nu(l,m) \exp[-2\pi i(u_{ij}(t)l \\ +v_{ij}(t)m)]dldm \qquad (5.15)$$

where

ν is the monochromatic frequency of the radiation,

(l, m) specify positions on the sky, they are direction cosines measured with respect to the u and v axes.

A_ν (l,m) is the normalized reception pattern of the antenna, assumed the same for all antennas.

I_ν (l,m) is the intensity distribution of the source.

At the VLA, the observed visibilities, generated by the array, and the true visibilities, can be related through (Perley, et al, 1988)

$$V_{ij}^{'}(t) = G_{ij}(t)V_{ij}(t) + \epsilon_{ij}(t) + \eta_{ij}(t), \qquad (5.16)$$

where

G_{ij} is the baseline-based complex gain,

ϵ_{ij} is the baseline-based complex offset,

η_{ij} is the complex noise.

The offset term is assumed to be negligible and the noise is assumed to be negligible after proper averaging of the data in the scan. Thus the observed and true visibilities are related through

$$V_{ij}'(t) = G_{ij}(t)V_{ij}(t) \qquad (5.17)$$

for each antenna pair (Perley, et al, 1988).

Since most data corruption occurs before the signal pairs are correlated, the baseline-based complex gain, G_{ij} can be approximated by the product of the two associated antenna-based complex gains, g_i and g_j, (Perley et al, 1988)

$$G_{ij}(t) = g_i(t)g_j^*(t)g_{ij}(t) = A_{ij}(t)\exp[i\Phi_{ij}(t)], \qquad (5.18)$$

where

$g_{i,j}(t)$ is the antenna-based complex gain for antenna i, j.

$g_{ij}(t)$ is the residual baseline-based complex gain, the "closure error".

$A_{ij}(t)$ is the antenna based amplitude correction, $A_{ij} = a_i(t)a_j(t)a_{ij}(t)$. $a_{ij}(t)$ is the closure error in amplitude.

$\Phi_{ij}(t)$ is the antenna based phase correction, $\Phi_{ij}(t) = \phi_i(t)-\phi_j(t)+\phi_{ij}(t)$. $\phi_{ij}(t)$ is the closure error in phase.

For a well designed system g_{ij} is within one percent of unity.

The goal of calibration is to estimate the true visibilities from the observed. The calibration consists of determining the complex gain of each antenna, as a function of time, from the observed visibilities. This is done through the use of unresolved point source calibrators. If the complex gain cannot be determined for any time region during the observations, data from this time frame, for these antenna, has to be discarded. If the observed visibility is represented by $V'_{ij} = A'_{ij}\exp(i\phi'_{ij})$ and the true by $V_{ij} = A_{ij}\exp(i\phi_{ij})$, then for a point-source calibrator of flux density S, $A_{ij} = S$, and $\phi_{ij} = 0$ and therefore:

$$V_{ij} = A_{ij}\exp(i\phi_{ij}) = S \qquad (5.19)$$

The calibration equations become

$$V'_{ij} = G_{ij}V_{ij} = A_{ij}(t)\exp\left[i\Phi_{ij}(t)\right]S = A'_{ij}\exp\left[i\Phi'_{ij}(t)\right] \tag{5.20}$$

which can be separated to give

$$A'_{ij} = a_i a_j a_{ij} S \tag{5.21}$$

and

$$\phi'_{ij} = \phi_i - \phi_j + \phi_{ij}. \tag{5.22}$$

At each time for which calibration data is available, this pair of equations exists for each of the N(N-1)/2 baselines. Thus the system of equations for the N antenna-based complex gains is highly over-determined and least squares can be used to find a_i and ϕ_i for all antennas, as functions of time. Observations of strong, unresolved continuum sources, of well known position and flux densities are used for this purpose. A primary flux calibrator is observed at least once during the observations and secondary phase calibrators are observed every 40 or so minutes to determine the antenna phase offsets.

Because the sensitivity is much greater in channel zero than in the line channels, calibration is carried out using this broadband channel and then applied to the line database. Briefly, the flux density of the primary calibrator, 3C48, is assigned its "known" value at the frequency of observation. Using the flux densities of the secondary calibrators as free parameters in the amplitude solutions, a solution for the amplitude and phase for each antenna is computed as a function of time. If any of the closure errors seem excessive (greater than approximately 10% or 10°) the data base is reinspected,

the offending data flagged and the solutions recalculated. Once acceptable solutions for the complex gains have been found using the primary calibration source, the fluxes of the secondary calibrators are determined, i.e., "bootstrapped" from the primary flux calibrator. The final step in the channel 0 calibration is to calibrate the flux of the galaxy from the calibrated phase calibrators. The calibration is applied to the galaxy dataset by a running mean, boxcar, interpolation of the amplitude and phase gains of the individual antennae (England, 1986). To calibrate the line data the calibration table is copied to the line data.

The final step in a spectral line calibration procedure is to calibrate the bandpass. This compensates for any variation of the antenna gain with frequency. Bandpass calibration is done by assuming a flat spectrum for the primary calibrator over the total spectral-line bandwidth, thus correcting for the complex gain variations across the spectral channels.

For both galaxies the primary flux calibrator was 3C48. For the 1991 C/D observations, the phase calibrators were 0237–233 for NGC1398 and 0445–221 and 0500+019 for NGC1784. In 1992, B/C observations, the calibrators were 0237–233 and 0406–180 for NGC1398 and 0500+019 for NGC1784

Neither galaxy displayed any serious problems with the data in either array configuration. All "shadowed" data, was flagged. Shadowing occurs when the signal going to one antenna is blocked or partially blocked by another antenna. This occurs for the shorter baselines when the object is low on the horizon and the projected baseline is shorter than the size of the antennae, 25m. In addition, there were one or two antennae that needed to be removed, either from the entire data set, or a large portion, due to mal-

functions. With the exceptions of these minor problems, good solutions were obtained fairly quickly and without excessive flagging of data.

Making and Cleaning Maps

The observed visibility is the Fourier transform (FT) of the modified sky brightness, I', the product of the true brightness and the single dish power pattern.

$$I^{'} = I(x,y)A(x,y) = \int\limits_{-\infty}^{+\infty} \int\limits_{-\infty}^{+\infty} V^{'}(u,v)\exp[2\pi i(ux+vy)]dudv. \tag{5.23}$$

This equation is used to obtain the source brightness distribution from the observed visibilities. The Fourier inversion of the observed visibility data, for each channel, gives a single dirty channel map, the HI distribution over a given velocity range.

Due to the large number of visibilities in a typical spectral-line data set, the only feasible way to evaluate the Fourier transform is to use a Fast Fourier Transform, FFT. However, to use a FFT, the visibilities values must lie on rectangular grid points, m × n, where m and n are usually integer powers of two. Actual visibilities are observed at discrete (u,v) points, and of course do not lie on such a grid. Thus the data must be averaged to estimate values at grid points from nearby observed points. This process is referred to as gridding. To produce the gridded visibilities, the observed visibilities are weighted, if desired, and convolved to produce a continuous function. This convolution is then resampled at the center of each grid cell. For both galaxies, square grids were used.

If the resampled, convolved visibilities are denoted, V^R, then:

$$V^R = R\Big(C * V^W\Big), \tag{5.24}$$

where V^W is the weighted, sampled visibility function, defined as WV′. Thus

$$V^R = R(C * (WV')),\qquad(5.25)$$

where R is the resampling function, C is the convolving function. The convolution theorem states that the Fourier transform of the product of two functions is the convolution of their FTs. The Fourier transform of V^R, FV^R, is the dirty image, I'_D.

$$I'_D = \hat{V}^R = \hat{R} * [(\hat{C})(\hat{W} * \hat{V}')].\qquad(5.26)$$

where ˆ represents the Fourier Transform (Perley et al, 1989).

To do the averaging, the data is convolved with the convolving function, C(u,v). An ideal convolving function has a Fourier transform that is flat across the image and then drops off rapidly. The choice of C is important because a good convolution function will helps suppress aliasing of sources outside the map onto the map. The C function recommended at the VLA is a separable spheroidal function:

$$C(u,v) = C(u)C(v)\qquad(5.27)$$

where

$$C(x) = |1 - \eta^2(x)|^\alpha \psi_{\alpha 0}(\pi m/2, \eta(x))$$
$$\eta(x) = \frac{2x}{m\Delta x}\qquad(5.28)$$

and $\psi_{\alpha 0}$ is a 0-order spheroidal function (England, 1986)

$$\psi_{\alpha,0} = \left(1 - \eta^2\right)^{-\frac{\alpha}{2}} S_{\alpha,\alpha}(C,\eta).\qquad(5.29)$$

S is a prolate spheroidal wave function. $\psi_{\alpha 0}$ is used at the VLA, with m = 6 and α = 1 or −1/2 being typical.

Once convolved, the map is resampled to produce the gridded values using a "bed of nails" function, given in term of Bracewell's two dimensional "sha" function, III, by

$$R(u,v) = III\left(\frac{u}{\Delta u}, \frac{v}{\Delta v}\right) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \delta\left(j - \frac{u}{\Delta u}, k - \frac{v}{\Delta v}\right). \qquad (5.30)$$

where Δu, Δv define the cell sizes, the separation between grid points.

Before making a map, the gridded data may be weighted to control the beam shape. Weighting may take the form of tapering or density weighing. Tapering multiples the weights of all points by a factor which decreases at increasing distances from the origin of the (u,v) plane. This underweights the longer baselines, in order to decrease the sidelobes of the array. These outer portions of the (u,v) plane are less densely filled with data, and thus less well determined. The low density of observed points in the outer (u,v) plane is directly responsible for the strong inner sidelobes of the beam and so tapering to reduce these baselines improves the sensitivity but does so at the expense of resolution, as the larger baselines are responsible for high resolution. The tapering function is usually separable, with the most common form being a Gaussian.

Density weighing can also be applied, in the form of either natural or uniform weighting. For natural weighting, all data points are treated equally and the weight of a cell is proportional to the number of data points contained within the cell. Because there are many more data points for short baselines, at the center of the u,v plane, natural weighting weighs the center heavily, giving the best signal to noise ratio, but lower resolution. Uniform weighting, on the other hand, assigns equal weight to all cells that contain data points. Thus, the under-sampled, longer baselines are emphasized, giving

higher resolution, but a lower signal to noise. For uniform weighing the beam is often specified largely by a tapering function.

Because the galaxies observed were rather weak, naturally weighted maps were made first. For both galaxies, this gave good results; at no point was the signal lost in the noise for either array. Experiments were made with uniform weighting but only natural weighted results will be included here.

Before the dirty maps were made, the map and cell sizes needed to be determined. The total map was made 3 to 4 times greater than the total expected galaxy size. In order to get good sampling the cell size was picked so as to make it three to four times smaller than the short axis of the synthesized beam. At this point equation (5.26) can be solved and the transform preformed producing a dirty beam and its dirty map for each line channel. The dirty maps represent the true brightness distribution convolved with the dirty beam. The dirty maps are then stored in a dirty cube, with axes of right ascension, declination and frequency. Figures 5-2 and 5-3 show the dirty beam and a dirty map for channel 16 for NGC 1398. Maps were made using the combined data set. The dirty beam is plotted only to the 10% level. There is significant response at the 1% level.

The total maps sizes are 256×256 cells, with each cell representing $7''$. From figure 5-2, the beam size at half maximum is approximately $20'' \times 22''$ with a position angle of $-35.9°$. At a distance of 16.1 Mpc (Tully, 1988) this corresponds to a linear resolution of 1.56 Kpc $\times$ 1.72 Kpc. In figure 5-3, the sidelobes of the dirty beam are apparent.

Center at RA 03 36 45.000 DEC -26 29 55.00
Peak flux = 1.0E+00 JY/BEAM
Levs = 1.0E-01 * (1.0, 3.0, 5.0, 7.0, 9.0)
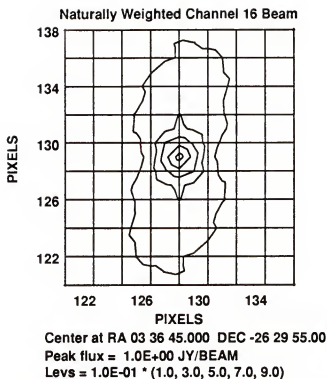
Figure 5-2: The dirty beam for channel 16, NGC 1398.

The next step in the image making process is to subtract out the continuum. Three or four line-free channels at the beginning and end of the data cube are averaged together to form a single mean continuum map. This map is then subtracted from all line channel maps. Figure 5-4 shows the continuum map made by combining five line free channels, for NGC 1398.

The channels with line emission were then CLEANED down to two times the continuum-free noise level using the channel 0 dirty beam for the individual data sets and the channel 16 beam for the combined data sets. Cleaning is necessary because there are unsampled regions in the (u,v) plane, especially at the longer baselines. Application of a FFT upon these regions gives an estimate of zero brightness and thus the dirty maps

NGC1398 IPOL 1413.711 MHZ NW1398.IMAP.16

Center at RA 03 36 44.739 DEC -26 29 58.50
Peak flux = 1.3507E-02 JY/BEAM
Levs = 9.0000E-04 * (-1.00, 1.000, 3.000,
6.000, 9.000, 12.00, 15.00)

Figure 5-3: The dirty channel 16 map for NGC 1398.

NGC1398 IPOL 1410.781 MHZ NW1398.SUM.1

Center at RA 03 36 44.739 DEC -26 29 58.50
Peak flux = 1.3776E-02 JY/BEAM
Levs = 9.000E-04 * (-1.0, 1.0, 3.0,6.0, 9.0, 12.0, 15.0)

Figure 5-4: A dirty continuum map for NGC 1398, made from 5 line free channels.

are produced with all unsampled visibilities set to zero. This gives rise to a distorted synthesized beam, and is responsible for the dirty beam's sidelobes. The channel maps of the source derived from the FT of the available amplitude and phase recordings suffer similar distortion. To overcome this problem, these dirty maps are cleaned. CLEAN is an iterative process to remove from the dirty map the distortion due to the dirty beam. To achieve this, all individual structures in the map are considered due to point sources and the radio source is the sum of these point sources. The brightest point source is found and the beam shape is subtracted from this most prominent maximum of the radio ma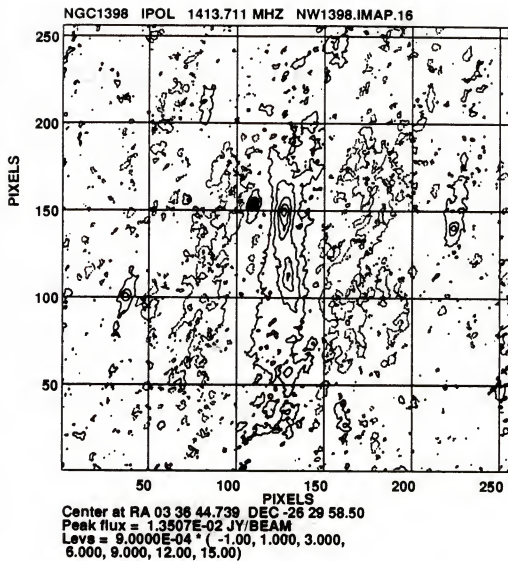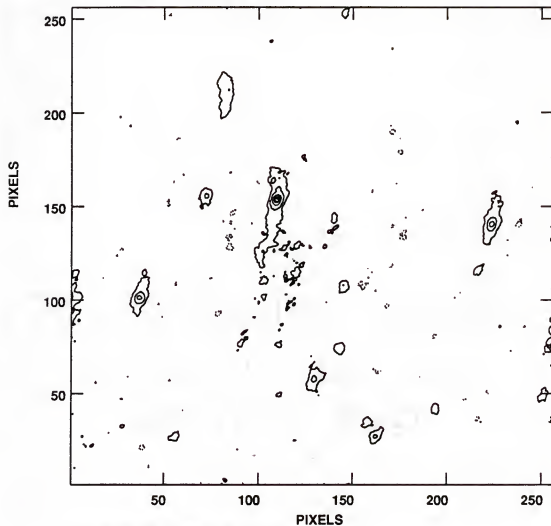p. Thus, the highest peak and its associated distorted sidelobes are removed. The same procedure is then successively applied to the remaining features on the map. Finally the map is restored by returning at appropriate amplitudes the various components with a "clean beam"; the beam that would have resulted from a completely filled aperture. The clean beam is an elliptical Gaussian function, specified by its major and minor axis, and the position angle on the sky of its major axis. These parameters are determined by fitting to the inner region of the dirty beam (Hey, 1983). The cleaned channel maps represent the signal with the dirty beam removed. Thus any blank sky in such maps should show only random noise and no sidelobe structure. The rms noise level should be close to the theoretical noise prediction. For naturally weighted maps the theoretical rms noise in mJy/beam for 21 cm observations is given by

$$S_{rms} = \frac{10}{\sqrt{N(N-1)n\Delta t\Delta f}} \tag{5.31}$$

where

N = the number of antennae, 26

n = the number of IFs, 2

Δt = the total observing time in hours, 10.5 for NGC1398 and 11 for NGC 1784.

Δf = the bandwidth in MHz, for the rms in one channel this is 1.22×0.0977MHz.

For the combined data sets, this gives values of 2.5×10^{-4} Jy/beam for both NGC1398 and NGC1784.

Table 5.1: rms noise levels for the combined data sets.

rms (Jy/beam)	NGC1398	NGC1784
Theoretical	2.5×10^{-4}	2.5×10^{-4}
Dirty line free maps (DLFM)	5.6×10^{-4}	5.0×10^{-4}
DLFM with continuum subtracted out	3.5×10^{-4}	3.4×10^{-4}
Clean maps	3.6×10^{-4}	3.4×10^{-4}

Figure 5-5 shows the cleaned channel maps for NGC 1398. Four channels are shown per page starting with channel 7. Each pixel is 7″. Only the CLEANED boxes are shown, with the channel number in the bottom left corner of each box. The rms uncertainty on these maps is 3.6×10^{-4} mJy per beam solid angle, or 0.51 K.

The final step is to make moment maps of the dirty image cube. The zeroth moment of the brightness temperature with respect to velocity, equation (5.11) gives the HI density distribution, the first moment gives the radial velocity contours (equation 5.13). The moments are determined by summing, at each right ascension, declination point

NGC1398   1411.562 MHZ   IPOL   NW1398.ICLN.1

Center at RA 03 36 45.261 DEC -26 30 5.50
Peak flux = 9.3705E-03 JY/BEAM
Levs = 9.0000E-04 * (1.000, 3.000, 5.000,
7.000, 9.000)
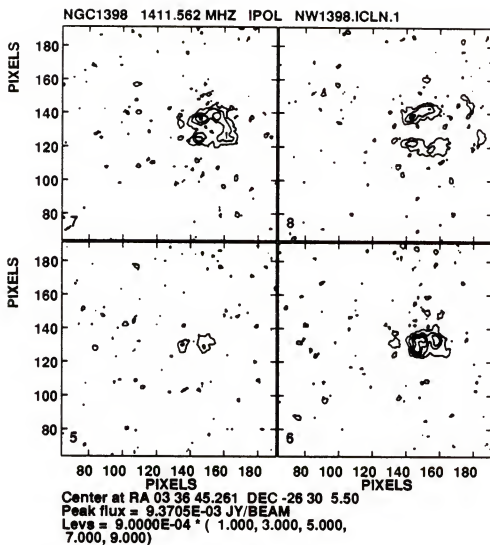
Figure 5-5: Cleaned channel maps for NGC 1398.
Channel numbers are given in the bottom left corner

NGC1398 1412.343 MHZ IPOL NW1398.ICLN.1

Center at RA 03 36 45.261 DEC -26 30 5.50
Peak flux = 9.3705E-03 JY/BEAM
Levs = 9.0000E-04 * (1.000, 3.000, 5.000,
7.000, 9.000)

Figure 5-5: continued

NGC1398 1413.125 MHZ IPOL NW1398.ICLN.1

Center at RA 03 36 45.261 DEC -26 30 5.50
Peak flux = 9.3705E-03 JY/BEAM
Levs = 9.0000E-04 * (1.000, 3.000, 5.000,
7.000, 9.000)

Figure 5-5: continued

NGC1398 1413.906 MHZ IPOL NW1398.ICLN.1

Center at RA 03 36 45.261 DEC -26 30 5.50
Peak flux = 9.3705E-03 JY/BEAM
Levs = 9.0000E-04 * (1.000, 3.000, 5.000,
7.000, 9.000)

Figure 5-5: continued

NGC1398   1414.687 MHZ   IPOL   NW1398.ICLN.1

Center at RA 03 36 45.261 DEC -26 30 5.50
Peak flux = 9.3705E-03 JY/BEAM
Levs = 9.0000E-04 * (1.000, 3.000, 5.000,
7.000, 9.000)

Figure 5-5: continued

of the dirty cube, over the third dimension, the frequency. The intensity is smoothed/averaged using Hanning smoothing for the velocity coordinate, and Gaussian for the spatial coordinate. Moment maps for the combined data sets for NGC 1398 are presented in figures 5-6 and 5-7. Figure 5-6 shows the zeroth moment map, the density distribution for NGC 1398. The gas is severely depleted in the central region. The peak flux corresponds to a column density of 6.1×10^{21} at/cm^2. Figure 5-7 shows the first moment map, the line of sight velocity contours for NGC 1398. The contours are fairly smooth, implying that there is not too much non-circular motion. The contours close, implying a falling rotation curve.

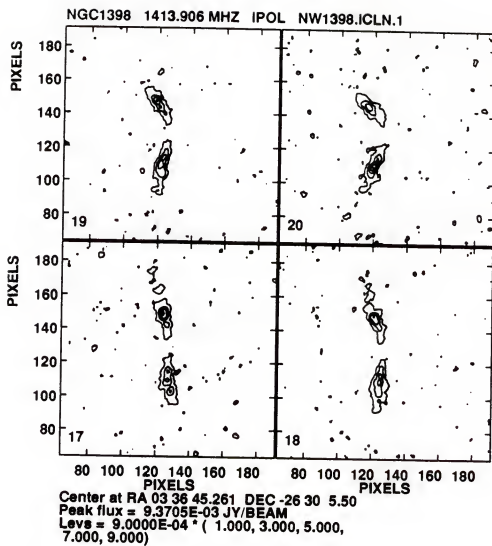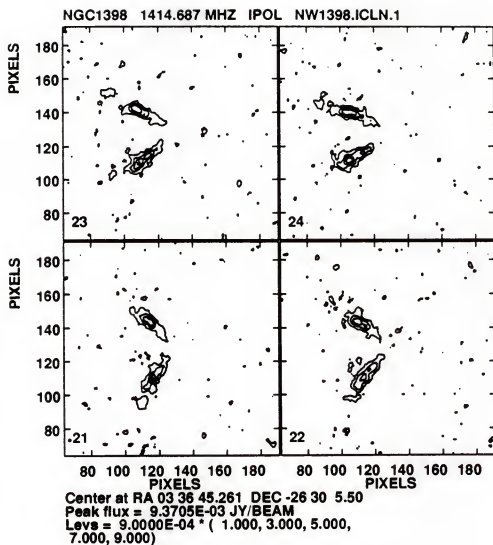Maps for NGC 1784 were made in the same way and are presented in figures 5-8 through 5-11. Figure 5-8 shows the dirty channel 16 beam. Each pixel is 7″. The resolution (FWHP) is 21″ × 28″, corresponding to a linear resolution of 2.92 × 3.4 Kpc at a distance of 28.7 Mpc (Tully, 1988). The beam position angle is 44°.

Cleaned maps for channels 8 through 23 are shown in figure 5-9. Figures 5-10 and 5-11 are the zeroth and first moment maps. The maximum column density is 1.84×10^{21} at/cm^{-3}. The density is not depleted in the center as it was for NGC 1398, but the first moment map indicates that either large non-circular motions are present or that the disk is warped.

In this chapter, the observations, calibration and map making for two barred spiral galaxies were discussed. Infrared observations exist for these galaxies and in the future, an attempt will be made to match the models presented in earlier chapters with the infrared surface photometry and the H I observations.

NGC1398   IPOL   NW1398.MOM0.2

Center at RA 03 36 43.436   DEC -26 29 41.00
Peak flux =   2.3560E+03 JY/B*M/S
Levs =   2.3560E+02 * (1.000, 2.000, 3.000,
  4.000)
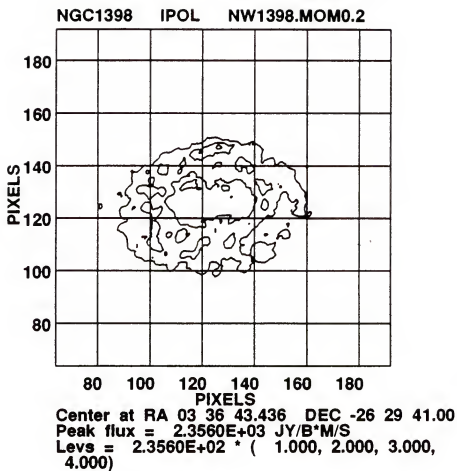
Figure 5-6: The zeroth moment map, the density distribution for NGC 1398.

NGC1398   IPOL   NW1398.MOM1.2

Center at RA 03 36 43.436   DEC -26 29 41.00
Peak flux =   1.9256E+06 M/S
Levs =   9.0965E+05 * (1.000, 1.100, 1.200,
  1.300, 1.400, 1.500, 1.600, 1.700, 1.800,
  1.900, 2.000, 2.060)
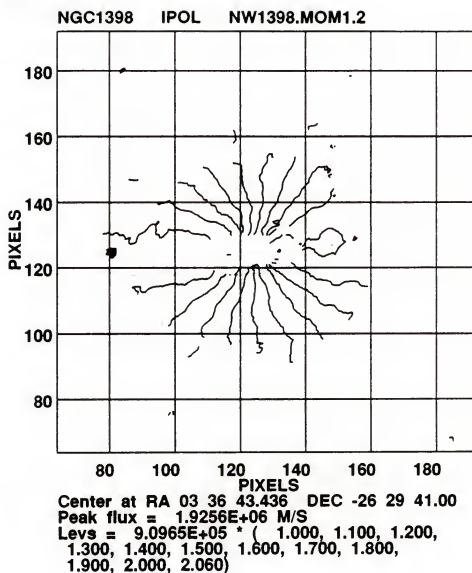
Figure 5-7: The first moment map, the radial velocity distribution for NGC 1398.

NGC1784   BEAM   1409.494 MHZ

Center at RA 05 03 6.600 DEC -11 56 26.50
Peak flux = 9.9999E-01 JY/BEAM
Levs = 9.9999E-02 * (0.10, 1.0, 3.0,5.0, 7.0, 9.0)
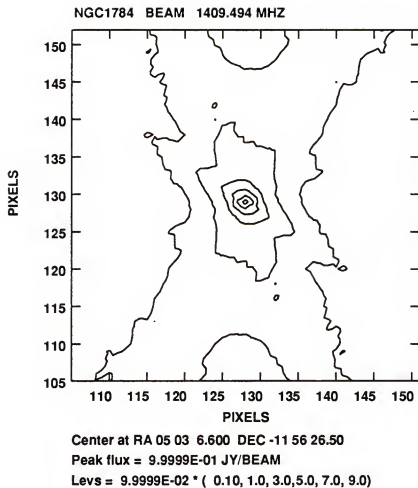
Figure 5-8: The dirty channel 16 beam for NGC
1784. The beam is shown down to the 1% level.
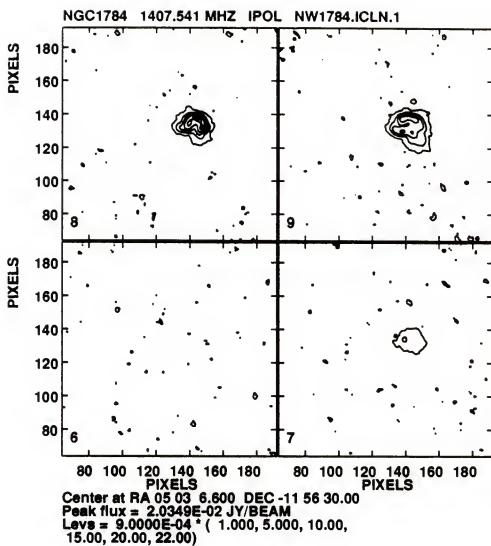
Figure 5-9: Cleaned channel maps for NGC 1784.
Channel numbers are given in the bottom left corner.

NGC1784   1408.322 MHZ   IPOL   NW1784.ICLN.1

Center at RA 05 03 6.600   DEC -11 56 30.00
Peak flux = 2.0349E-02 JY/BEAM
Levs = 9.0000E-04 * (1.000, 5.000, 10.00,
15.00, 20.00, 22.00)
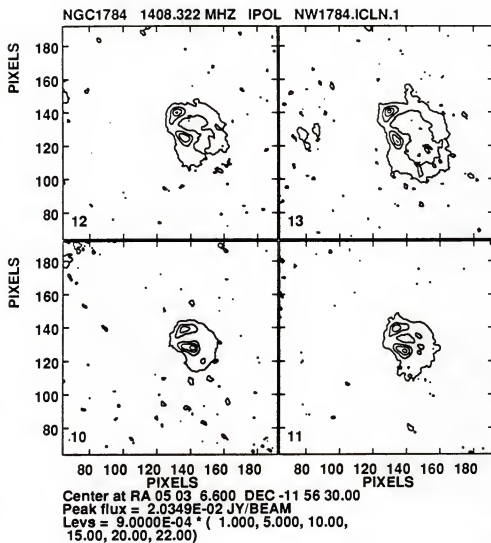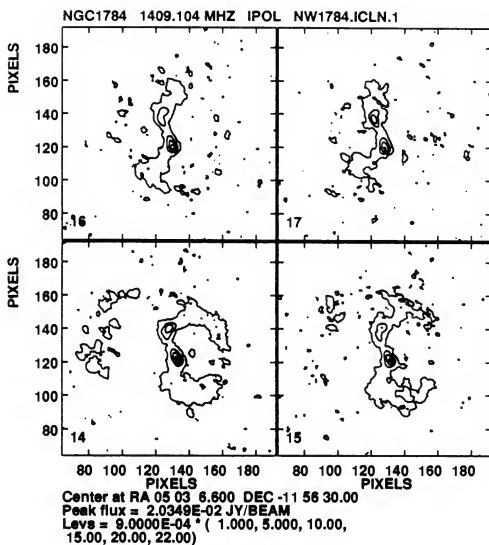
Figure 5-9: continued

Figure 5-9: continued

NGC1784   1408.322 MHZ   IPOL   NW1784.ICLN.1

Center at RA 05 03 6.600 DEC -11 56 30.00
Peak flux = 2.0349E-02 JY/BEAM
Levs = 9.0000E-04 * (1.000, 5.000, 10.00,
  15.00, 20.00, 22.00)

Figure 5-9: continued

NGC1784   1410.666 MHZ   IPOL   NW1784.ICLN.1

Center at RA 05 03 6.600 DEC -11 56 30.00
Peak flux = 2.0349E-02 JY/BEAM
Levs = 9.0000E-04 * (1.000, 5.000, 10.00,
15.00, 20.00, 22.00)
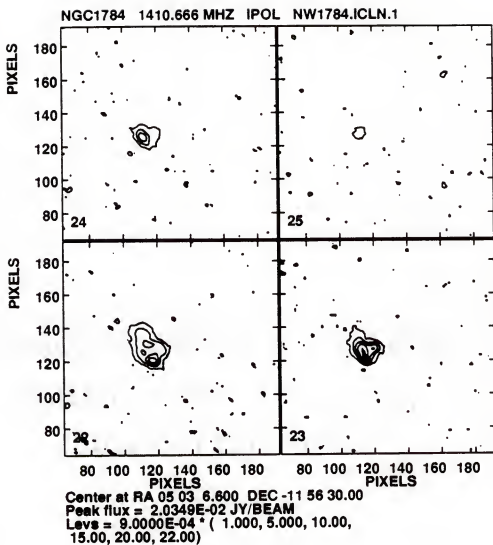
Figure 5-9: continued
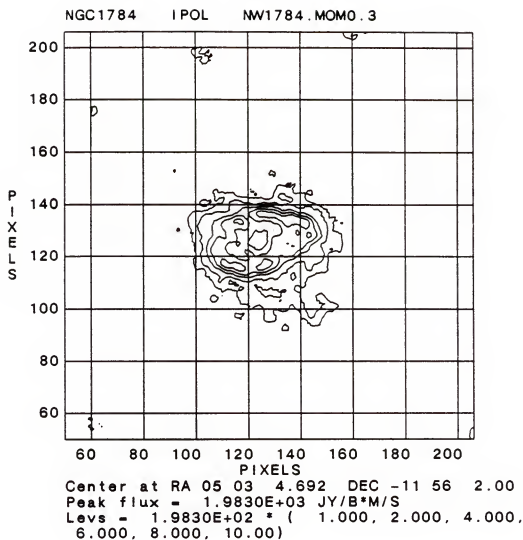
Figure 5-10: The zeroth moment map, the density distribution for NGC 1784.

NGC1784   IPOL   NW1784.MOM1.3

Center at RA 05 03 4.692   DEC -11 56 2.00
Peak flux =   2.8937E+06 M/S
Levs =   2.4596E+06 * (0.900, 0.920, 0.940,
 0.960, 0.980, 1.000, 1.020, 1.040, 1.060,
 1.080, 1.100, 1.120, 1.140, 1.160)

Figure 5-11: The first moment map, the radial velocity distribution for NGC 1784.

CHAPTER 6
SUMMARY AND CONCLUSIONS

In the preceding chapters the development and testing of a two-component code has been described. In Chapter 2, the algorithm was described and shown to be trustworthy. In Chapter 3, a method of calculating the dispersional velocities was presented. It was shown that either the addition of a halo or large noncircular motions were required to stabilize a cold disk. The halo mass necessary to stabilize the disk against bar formation is approximately 2 times the disk mass. Such a halo decreased the ratio of the peculiar velocity to the rotational velocity to a value close to that observed in our galaxy. If heating alone is used to stabilize the disk, the noncircular motions are well in excess of those observed. If both a halo and heating are included, with the halo mass comparable to the disk mass, the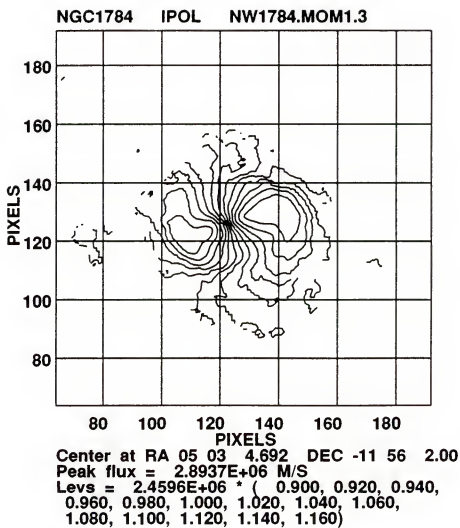 model will stabilize with random velocities (compared to the rotational velocity) of about twice the value observed. In Chapter 4, methods of building galactic models with structure were presented. The first method, adding a halo to a cold two component disk, was developed following the work of Carlberg and Freeman, CF, (1985). This provided a further test of the algorithm. Although the two codes are different, CF calculate the potential on a polar grid, and use a sticky particle gas scheme, their results were duplicated. Further models were run to explore the effects of varying the gas parameters. This is important as the gas scheme employed has not been widely used. Finally, a second method of developing bar and spiral structure was presented. In this procedure, the stabilizing stellar peculiar velocities were decreased in

the center of the model while maintained in the outer disk. This caused a bar to form in the interior, surrounded by a stable stellar disk. In the gas component, a bar and spiral structure developed. By varying the size of the unstable regions, the rate at which k varies, and the halo mass, a wide variety of spiral structure could be developed.

This rest of the chapter will be used to discuss some of the possible future applications of the numerical tool that has been developed. The answer is twofold; to use the code in an attempt to model actual barred spiral galaxies and, secondly, to use it to make theoretical studies.

It is important to compare theoretical findings with observations. A model may make a prediction that observations either prove or disprove. In either case, something is learned. Or observations may exist for which there is little explanation or physical understanding. Models may then be used to make sense of the observations. It is assumed that a model that more closely fits the observations is closer to mimicking the physical reality of an actual galaxy. It is worth studying if and how model "observations" change in a systematic way when a model parameter changes. If they do and one case of the parameter is in better agreement with observation, then, provided that the physical meaning of the parameter is known, something will have been learned about the physics of the actual galaxy. This provision should be kept in mind when attempting to model any one particular galaxy. It is perhaps more important to study general trends that occur (in the pseudo observations) when a parameter is changed. Systematic changes that occur in the model when parameters are changed are worth studying and trying to compare to observations. But, for galaxy models at least, the codes are still in a crude

phase and detailed matching to observations may not be possible. As Miller (Miller, Prendergast and Quirk, 1970) notes: "As with any gravitational n-body calculation, only qualitative features of this one are to be trusted."

VLA radio observations provide two main pieces of information for each galaxy, a H I density contour map and an observed radial velocity contour map. Infrared surface photometry provided information about the stellar mass distribution, especially in the bar. In order to compare simulated models to 21 cm HI and infrared observations, the models must be projected at the observed galaxy's inclination, the angle between the galaxy and the plane of the sky. Following the notation of Mihalas and Binney, let (R,θ) be polar coordinates in the plane of the galaxy, and (ρ,ϕ) the corresponding quantities in the plane of the sky. In order to compare a model density with VLA observations, each particle in the galaxy in projected into the plane of the sky via;

$$R^2 = \rho^2 \left(\cos^2 \phi + \sec^2 i \sin^2 \phi \right) \tag{6.1}$$

and

$$\tan \theta = \sec i \tan \phi \tag{6.2}$$

Particles are then binned in the sky plane on a grid the same size as that used to produce the VLA maps. The density in each bin is calculated and convolved with a two dimensional gaussian beam, the same size as that produced by the VLA. A sample of this procedure is shown for a typical density plot. The upper half of figure 6-1, shows a model result that has been projected by 50° and convolved with a 22″ × 22″ gaussian beam.

The procedure to obtain the observed radial velocities is similar. First, the observed radial velocity in the plane of the sky is calculated from the model radial and azimuthal

velocity in the plane of the galaxy, i.e.,

$$V_r(\rho, \phi) = v_{sys} + \Pi(R, \theta) \sin \theta \sin i + \Theta(R, \theta) \cos \theta \sin i \qquad (6.3)$$

where v_{sys} is the systemic velocity, Π is the radial velocity and Θ, the azimuthal velocity in the plane of the galaxy. Each particle is projected into the plane of the sky as before and binned in the plane of the sky. The radial velocities in one sky bin are averaged. The velocity map is density weighted and convolved with a gaussian beam. Examples of the velocity contour maps are shown on the bottom half of figure 6-1. The effect of the bar is seen clearly in the velocity contours. In the future, it is hoped that comparing model observations such as these to actual observations, such as those presented in Chapter 5, will prove to be fruitful.

Throughout the work a number of the limits of the method and the code have been pointed out as they arise. Many of these provide useful starting points for future work. The option of a self-consistent halo provides an intriguing possibility. In addition, the Elmegreens (Elmegreen and Elmegreen, 1990) believe it is possible to determine the location of many of the major resonances from optical features such as kinks, bifurcations, and spurs in the spiral arms as well as the end of star-formation ridges. Very little work has been done so far to locate the resonances in the models. It would be fruitful to attempt this and to compare the results with the observational and theoretical (orbit calculations) results.
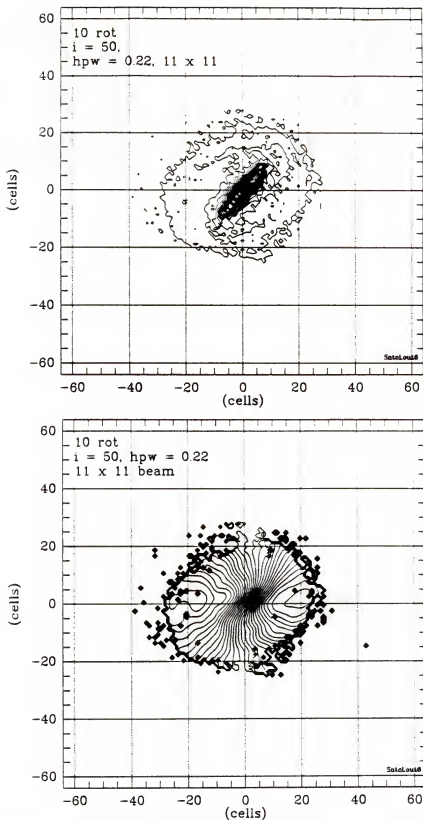
Figure 6-1: Pseudo-observations made from a model showing the projected density (top) and radial velocity curves (bottom) convolved with a gaussian beam.

In addition, it seems that the most important feature of the gas is that it be dissipative and that the exact form of the dissipation is unimportant. This is fortunate as the real state of the gas in the ISM is not well known. However, this aside, the gas component of the models could be made more sophisticated, in particular, it would be interesting to allow the gas clouds to evolve into stars and then to allow the stars to supernovae and return to the gas component. This may also help to maintain stellar spiral structure. One of the inconsistent findings of almost all numerical simulations thus far is that the stellar heating is well in excess of that observed. This causes any spiral structure that may appear early in the models to be smeared out. Sellwood and Carlberg (1984) show that accretion can be used to effectively cool the stellar disk and help maintain stellar spiral structure. This would be worth adding to the present code.

Finally, the start-up conditions of these models are not terribly unrealistic but no attempt is made to claim that actual galaxies begin in this configuration. The 3-dimensional problem of the collapse of the galaxy is quite different from the one explored within but it would be rewarding to make the code 3-dimensional and try to solve the problem from the beginning.

REFERENCES

Ball, R. 1984. Ph. D. dissertation, U. of Florida, Gainesville.

Bardeen, J.M. 1975. In *I. A. U. Symposium 69, Dynamics of Stellar Systems*, ed. Hayli, A., Reidel Dordrecht, the Netherlands.

Bertin, G. 1990. In *Galactic Models*, ed. Buchler, J. R, Gottesman, S. T. and Hunter, J. H., New York Academy of Sciences, New York.

Binney, J. J., and Tremaine, S. 1987. *Galactic Dynamics*, Princeton University Press, Princeton, N.J.

Burton, W. B. and Gordon, M. A. 1978. Astron. Astrophys. **63**, 7.

Carlberg, R. G and Freedman, W. L. 1985. Ap. J. **298**, 486.

Chandrasekhar, S. 1942. *Principles of Stellar Dynamics.* University of Chicago Press. Reissued by Dover 1960.

Contopoulos, G. and Grosbøl, P. 1986. Astron. Astrophys. **155**, 11.

Contopoulos, G. and Grosbøl, P. 1988. Astron. Astrophys. **197**, 83.

Cooley, J. W. and Tukey, J. W. 1965. Math. Comp. **19**, 297.

Elmegreen, B. G. 1991. In *I. A. U. Symposium 146, Dynamics of Galaxies and Their Molecular Cloud Distributions*, eds. Combes, F. and Casoli, F. Reidel Dordrecht, the Netherlands.

Elmegreen, D. M. and Elmegreen, B. G. 1990. Ap. J. **355**, 52

England, M. N. 1986. Ph. D. dissertation. University of Florida, Gainesville.

Efstathiou, G., Lake, G. and Negroponte, J. 1982. M.N.R.A.S. **199**, 1069.

Fujimoto, M. 1968. In *I. A. U. Symposium 29, Non-Stable Phenomena in Galaxies* Arnebuab Academy of Sciences, Yerevan.

Gingold, R. A. and Monaghan, J. J. 1977. M.N.R.A.S. **181**, 375.

Goldreich, P. and Lynden-Bell, D. 1965. M.N.R.A.S. **130**, 125.

Hey, J. S. 1983. *The Radio Universe*, Pergamon Press, Oxford.

Hockney, R. W. and Hohl, F. 1969. A. J. **74**, 1002

Hockney, R. W. and Brownrigg, D. R. K. 1974. M.N.R.A.S. **167**, 351.

Hohl, F. 1971. Ap. J. **168**, 343.

Hohl, F. 1972. J. Comp. Phys. **9**, 10

Hohl, F. 1975. In *I. A. U. Symposium 69, Dynamics of Stellar Systems*, ed. Hayli, A., Reidel Dordrecht, the Netherlands.

Hohl, F. 1978. A. J. **83**, 768.

Hohl, F. and Hockney, R.W., 1969. J. Comp. Phys. **4**, 306.

Hunter, C. 1963. M.N.R.A.S. **126**, 299.

Hunter, J. H., Ball, R., and Gottesman, S. T. 1984. M.N.R.A.S. **208**, 1.

Jeans, J. H. 1919. Phil. Trans. Roy. Soc. London A, **218**, 157.

Julian, W. H. and Toomre, A. 1966. Ap. J. **146**, 810.

Lewis and Freeman, A. J. **97**, 139.

Lin, C. C. and Shu, F. H. 1964. Ap. J. **140**, 646.

Lucy, L. B. 1977. A. J. **82**, 1013.

Mihalas, D. and Binney, J. 1981. *Galactic Astronomy*, Second Ed., Freeman and Co., San Francisco.

Mihalas, D. and Routly, P. M. 1968. *Galactic Astronomy*, Freeman and Co., San Francisco.

Miller, R. H. 1976. J. Comp. Phys. **21**, 400.

Miller, R. H. and Prendergast, K. H. 1968. Ap. J. **151**, 699.

Miller, R. H., Prendergast, K. H. and Quirk, W. J. 1970. Ap. J. **161**, 903.

Miller, R. H. and Smith, B. F. 1979. Ap. J. **227**, 785.

Monaghan, J. J. 1985. Computer Review Reports **3**, 71.

Ostriker, J. P. and Peebles, P. J. E. 1973. Ap. J. **186**, 467

Perley, R. A., Schwab, F. R. and Bridle, A. H. eds. *Synthesis Imaging in Radio Astronomy*, 1989. Astronomical Society of the Pacific Conference Series, San Francisco.

Persic, M. and Salucci, P. 1990. M.N.R.A.S. **245**, 577.

Press, W., Flannery, B., Teukolsky, S. and Vetterling, W. 1989. *Numerical Recipes*, Cambridge University Press, Cambridge

Roberts, W. W. 1969. Ap. J. **158**, 123.

Roberts, W. W., Roberts, M. S. and Shu, F. H. 1975. Ap. J. **196**, 381.

Roberts, W. W., Huntley, J. M. and van Albada, G. D. 1979. Ap. J. **233**, 67.

Rybicki, G. B. 1972. In *I. A. U. Colloquium 10, Gravitational N-Body Problem* ed. Lecar, M., Reidel Dordrecht, the Netherlands.

Sandage, A. and Tammann, G. A. 1981. *A Revised Shapley-Ames Catalog of Bright*

Galaxies. Carnegie Institution, Washington, D.C.

Sanders, R. H. 1977. Ap. J. **216**, 916.

Sanders, R. H. and Prendergast, K. H., 1974. Ap. J. **188**, 489.

Sanders, R. H. and Huntley, J.H. 1976. Ap. J. **209**, 53.

Schussler, M. and Schmitt, D. 1981. Astron. Astrophys., **97**, 373.

Schwarz, M. P. 1981. Ap. J. **247**, 77

Sellwood, J. A. 1980. Astron. Astrophys. **89**, 296.

Sellwood, J. A. 1981. Astron. Astrophys. **99**, 362.

Sellwood, J. A. 1983. J. Comp. Phys. **50**, 337.

Sellwood, J. A. 1985. M.N.R.A.S. **217**, 127.

Sellwood, J. A. 1987. Ann. Rev. Astron. Astrophys. **25**, 151.

Sellwood, J. A. 1990. In *Galactic Models*, ed. Buchler, J. R., Gottesman, S. T. and Hunter, J. H. New York Academy of Sciences, New York.

Sellwood, J. A. and Carlberg, R. G. 1984. Ap. J. **282**, 61.

Shu, F. J 1982. *The Physical Universe.* University Science Books, Mill Valley, California.

Toomre, A. 1963. Ap. J. **138**, 385.

Toomre, A. 1964. Ap. J. **139**, 1217.

Toomre, A. 1969. Ap. J. **158**, 899.

Toomre, A. 1977. Ann. Rev. Astron. Astrophys. **15**, 437.

Toomre, A. 1981. In *The Structure and Evolution of Normal Galaxies*, ed. Fall and Lynden-Bell, Cambridge University Press, Cambridge.
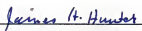
Tully, R. B. 1988. *Nearby Galaxies Catalog*, Cambridge University Press, Cambridge.

Wielen, R. 1974. P. A. S. P. **86**, 341.

BIOGRAPHICAL SKETCH

E. Moore was born in New York in 1963. She attended Earl L. Vandermeulen High School and then went on the Williams College in Massachussets. She joined the cross-country team and quickly became addicted to long distance running through the Berkshire mountains. The addiction has continued, despite the change in scenery. She took the second half of her sophomore year off to learn to cook in a small health food store/bakery. Her junior year was spent at the University of Sussex in Falmer, England. During the especially cold winter of her senior year, Florida seemed a wise choice of graduate schools, and so in the fall of 1985, she entered the Ph.D. program at the University of Florida in astronomy. She received her master's degree in 1988. Now, seven long years after she entered graduate school, she is looking foward to leaving.
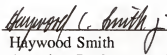
I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

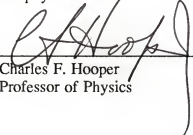James H. Hunter, Chair
Professor of Astronomy

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Stephen T. Gottesman, Cochair
Professor of Astronomy

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Haywood Smith
Associate Professor of Astronomy

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Charles F. Hooper
Professor of Physics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Henry E. Krandrup
Professor of Astronomy

This dissertation was submitted to the Graduate Faculty of the Department of Astronomy in the College of Liberal Arts and Sciences, and to the Graduate School and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

May 1992

Dean, Graduate School